

UNIVERSIDAD AUTÓNOMA DE MADRID

Facultad de Ciencias

Departamento de Biología Molecular



**ROLE OF THE TRANSCRIPTION FACTOR TCFL5
ISOFORMS IN COLON CANCER TUMORAL PROCESSES,
PLURIPOTENCY AND DEVELOPMENT**

PAPEL DE LAS ISOFORMAS DEL FACTOR DE TRANSCRIPCIÓN TCFL5
EN PROCESOS TUMORALES DE CÁNCER DE COLON, PLURIPOTENCIA Y
DESARROLLO.

Tesis Doctoral

JAVIER GALÁN MARTÍNEZ

Madrid, 2019

UNIVERSIDAD AUTÓNOMA DE MADRID

Facultad de Ciencias

Departamento de Biología Molecular



ROLE OF THE TRANSCRIPTION FACTOR TCFL5 ISOFORMS IN COLON CANCER TUMORAL PROCESSES, PLURIPOTENCY AND DEVELOPMENT

**PAPEL DE LAS ISOFORMAS DEL FACTOR DE TRANSCRIPCIÓN TCFL5
EN PROCESOS TUMORALES DE CÁNCER DE COLON, PLURIPOTENCIA Y
DESARROLLO.**

Memoria presentada por el Licenciado en Bioquímica **Javier Galán Martínez** para optar al
grado de Doctor en Biomedicina.

Directores de tesis:

Dr. Manuel Fresno Escudero

Dra. Núria Gironès Pujol

Centro de Biología Molecular “Severo Ochoa”

Madrid, 2019

Esta memoria presentada por **D. Javier Galán Martínez** para optar al grado de Doctor en Biociencias Moleculares por la Universidad Autónoma de Madrid ha sido realizada en el Laboratorio de Activación del Sistema Inmune dirigido por el Dr. Manuel Fresno y la Dra. Núria Gironès en el Centro de Biología Molecular “Severo Ochoa”.

El Doctorando ha disfrutado de una beca pre-doctoral del Ministerio de Economía y Competitividad (BES-2014-070038). Este trabajo ha sido financiado por la ayuda a Proyecto de Investigación “*Prostanoides y Receptores Tipo Toll como Mediadores Clave y Potenciales Dianas Terapéuticas Enfermedades Inflamatorias Crónicas*” (SAF2013-42850-R) otorgado por el Ministerio de Economía y Competitividad.

Abreviaturas.....	15
Resumen / Abstract.....	23
Introducción.....	29
1. Factores de transcripción.....	29
1.1. Clasificación de los factores de transcripción.....	29
1.2. Basic Helix Loop Helix.....	30
1.3. Control de la actividad de los factores de transcripción.....	31
2. Cáncer.....	32
2.1. Factores de transcripción en cáncer.....	33
2.2. Desregulación de la actividad de los factores de transcripción en cáncer.....	34
2.3. Los factores de transcripción como diana terapéutica contra el cáncer.....	35
3. Cáncer de colon.....	36
3.1. Tipos de cáncer de colon.....	37
3.2. Factores de transcripción y cáncer de colon.....	38
4. Desarrollo embrionario temprano.....	38
4.1. Factores de transcripción en etapas tempranas del desarrollo embrionario.....	39
4.2. Los factores de transcripción OCT4, NANOG y SOX2 durante el desarrollo embrionario temprano.....	39
5. Pluripotencia.....	40
5.1. Factores de transcripción en pluripotencia.....	41
6. Isoformas y <i>splicing</i> alternativo.....	42
7. <i>Transcription Factor Like 5</i> (TCFL5).....	43
7.1. Funciones de las isoformas TCFL5 y CHA.....	43
7.2. TCFL5 en cáncer.....	45
7.3. TCFL5 en pluripotencia y desarrollo embrionario.....	45
Objetivos.....	49
Materiales y Métodos.....	53
1. Materiales.....	53
1.1. Líneas y cultivos celulares.....	53
1.2. Cultivo de embriones <i>in vitro</i>	54
1.3. Plásmidos.....	57
1.4. Soluciones.....	57
1.5. Bases de datos.....	58
1.6. Anticuerpos.....	59
1.7. Oligonucleótidos.....	59
2. Métodos.....	61
2.1. Alineamiento de secuencias.....	61
2.2. Transfección transitoria.....	61
2.3. Clonaje y crecimiento bacteriano.....	62

2.4. Generación de líneas estables por lentivirus.....	62
2.5. Generación de líneas estables por edición génica (<i>Self-cloning</i> CRISPR-Cas9).....	63
2.6. Diferenciación de mECS a línea germinal.....	64
2.7. Obtención y análisis de ADN.....	64
2.8. Obtención y análisis de ARN.....	66
2.9. Obtención y análisis de proteínas.....	67
2.10. Ensayos de luminometría.....	69
2.11. Proliferación.....	70
2.12. Formación de colonias.....	70
2.13. Migración por cerrado de herida.....	70
2.14. Tamaño y número de MCTS.....	71
2.15. Generación de tumores xenoinjertados.....	71
2.16. Ética.....	71
2.17. Estadística.....	72
Resultados.....	77
1. Descripción y expresión tisular en condiciones fisiológicas de las isoformas del factor de transcripción <i>TCFL5</i> en humano y de ratón.....	77
1.1. <i>TCFL5</i> se expresa de forma ubicua en humano principalmente en testículo y cerebro.....	77
1.2. <i>TCFL5</i> presenta varias isoformas en humano.....	78
1.3. Las isoformas de <i>TCFL5</i> se expresan en todos los órganos humanos.....	79
1.4. La expresión de <i>TCFL5</i> se encuentra bajo el control de diversas vías de señalización.....	81
1.5. Expresión de <i>Tcfl5</i> en ratón.	83
1.6. <i>Tcfl5</i> presenta varias isoformas en ratón.....	83
2. Análisis de la expresión en cáncer de colon y caracterización de la función específica de las isoformas <i>TCFL5</i> y <i>CHA</i> en procesos tumorales.....	85
2.1. La expresión de <i>TCFL5</i> se encuentra inducida en cáncer de colon.....	86
2.2. Efecto de la sobreexpresión de <i>TCFL5</i> y <i>CHA</i> en el fenotipo tumoral de líneas de cáncer de colon.....	88
2.3. <i>TCFL5</i> y <i>CHA</i> regulan su expresión de manera recíproca.....	92
2.4. Efecto de la deficiencia de <i>TCFL5</i> y <i>CHA</i> en el fenotipo tumoral de líneas de cáncer de colon.....	93
2.5. Análisis de los genes diana y proteínas de interacción de las isoformas <i>TCFL5</i> y <i>CHA</i>	97
3. Papel de las isoformas <i>TCFL5</i> y <i>CHA</i> en la regulación de genes de pluripotencia... 101	
3.1. <i>TCFL5</i> y <i>CHA</i> se encuentran inducidos en los MCTS y controlan su formación.....	101

3.2. TCFL5 y CHA no producen cambios significativos en los marcadores de CSCs.....	103
3.3. TCFL5 y CHA son represores de la expresión de genes de pluripotencia.	104
3.4. TCFL5 y CHA modifican la expresión génica en los tumores xenoinjertados.....	107
3.5. TCFL5 reprime la actividad de SOX2, NANOG y KLF4.....	107
3.6. TCFL5 y CHA se unen al promotor de <i>SOX2</i> y <i>KLF4</i>	109
3.7. TCFL5 y CHA reprimen la actividad del promotor de <i>SOX2</i>	110
4. Expresión de TCFL5 en etapas tempranas del desarrollo embrionario de ratón y su papel en diferenciación.....	112
4.1. TCFL5 se expresa en etapas tempranas del desarrollo embrionario y en células embrionarias de ratón.....	112
4.2. La deficiencia en TCFL5 modifica la expresión de SOX2 y NANOG en células embrionaria de ratón.....	114
4.3. La expresión de <i>Tcfl5</i> se induce durante la diferenciación de células primordiales germinales (PGCs).....	116
4.4. La deficiencia de TCFL5 inhibe la diferenciación a PGCs.....	117
Discusión	125
1. Expresión de TCFL5 en tejidos humanos y de ratón.....	125
2. Expresión de TCFL5 en cáncer de colon.....	127
3. Auto-regulación de TCFL5 y CHA.....	127
4. Papel diferencial en el fenotipo celular producido por las isoformas TCFL5 y CHA.	128
5. Interacción de TCFL5 con NFkB2 y BCL2.....	131
6. TCFL5 y CHA en la vía de WNT y NOTCH1.....	132
7. TCFL5 y CHA en CSCs y pluripotencia en líneas celulares de cáncer de colon.....	133
8. Expresión de TCFL5 durante el desarrollo embrionario temprano.....	135
9. TCFL5 en diferenciación a línea germinal.....	136
Conclusiones / Conclusions	143
Bibliografía	149
Anexo I: Figuras y tablas	165
Anexo II: Artículos	175

Abreviaturas

Abreviaturas

ADN	Ácido desoxirribonucleico
ADNc	Ácido desoxirribonucleico complementario
AKT	Serina/treonina quinasa o proteína quinasa B
ALDH1	Aldehído Deshidrogenasa 1
APC	Poliposis adenomatosa coli (siglas en inglés)
ARN	Ácido Ribonucleico
ARNm	Ácido Ribonucleico mensajero
ARNT	Translocador nuclear del receptor de hidrocarburos arílicos (siglas en inglés)
B-ALL	Leucemia linfoblástica aguda de células B (siglas en inglés)
BAX	Proteína X reguladora de apoptosis asociada a BCL2 (siglas en inglés)
BCA	Ácido bicinchonínico (siglas en inglés)
BCL2	Proteína de linfoma de célula B 2 (siglas en inglés)
bFGF	Factor de crecimiento fibroblástico básico (siglas en inglés)
bHLH	basic helix loop helix (siglas en inglés)
BMP4	Proteína morfogénica ósea 4 (siglas en inglés)
BSA	Albúmina de Suero Bovino (siglas en inglés)
CBFB	Subunidad Beta del Factor de Unión al Núcleo (siglas en inglés)
CD133	Prominina 1
CD2	Cluster de diferenciación 2 o Antígeno de superficie de células T
CDK	Quinasa dependiente de ciclinas (siglas en inglés)
CDX2	Factor de transcripción caudal tipo homeobox 2 (siglas en inglés)
CHIP	Inmunoprecipitación de cromatina (siglas en inglés)
CHIP-Seq	Secuenciación de inmunoprecipitación de cromatina (siglas en inglés)
Chr	Cromosoma
CLGN	Calmegina
CMN	Cobertura Media Normalizada
CMV	Citomegalovirus
COE	Coiller/Olf1/EBF
CRISPR	Repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas
CSC	Célula madre tumoral (siglas en inglés)
Ct	Umbral de ciclo
Dapi	4,6-diamino-2-fenilindol
DAPT	(2S)-N-[(3,5-Difluorofenil)acetil]-L-alanil-2-phenil]glicina 1,1-dimetiletil ester
Dazl	Delecionado en Azoospermia
DDB1	Proteína de unión al DNA específica de daño 1 (siglas en inglés)
DDC	Dopa Decarboxilasa
DDX51	Helicasa DEAD-box 51 (siglas en inglés)

DMEM	Medio Dulbecco Modificado de Egel (siglas en inglés)
DMSO	Dimetilsulfóxido
DN	Dominante negativo
DUX4	Doble homeobox 4
E47	Factor de transcripción 3 o TCF-3
EBF	Factor de células B temprano (siglas en inglés)
EDTA	Ácido etilen-diamino-tetra-acético
EGF	Factor de crecimiento epidérmico (siglas en inglés)
EGTA	Ácido tri-etilen-glicodiamino-tetra-acético
EPHB2	Receptor de epinefrina tipo 2 (siglas en inglés)
EPI	Epiblasto
EV	Vector vacío (siglas en inglés)
FBS	Suero Fetal Bovino (siglas en inglés)
FLAG	Péptido de reconocimiento DYKDDDDK
GAPDH	Gliceraldehído-3-fosfato deshidrogenasa (siglas en inglés)
GATA6	Proteína de unión GATA 6
GP130	Glicoproteína 130
GSK3 β	Glucógeno Sintasa Quinasa Beta 3 (siglas en inglés)
HATs	Histonas acetil transferasas
HDACs	Histonas deacetilasas
HEPES	Ácido tetra-dihidroxietil-peperazino-etanosulfónico
HEY	Hairy/Enhancer-Of-Split relacionado con la proteína motivo YRPW
HIF	Factor inducible por hipoxia (siglas en inglés)
HMG	Grupo de alta movilidad (siglas en inglés)
HPR	Enzima peroxidasa de rábano (siglas en inglés)
HSP	Proteína de choque de calor (siglas en inglés)
ICM	Masa celular interna (siglas en inglés)
ICN	Notch intracelular (siglas en inglés)
ID	Proteína Inhibidoras de unión al ADN (siglas en inglés)
IF	Inmunofluorescencia
IMDM	Medio Iscove Modificado de Dulbecco
IP	Inmunoprecipitación
iPS	Células inducidas pluripotentes (siglas en inglés)
JAK	Quinasa Janus (siglas en inglés)
KLF	Factor tipo Kruppel (siglas en inglés)
KO	Knock-Out
KRAS	Homólogo del oncogen viral Kirsten Rat Sarcoma
LGR5	Receptor acoplado a proteína G 5 con repeticiones ricas en leucinas
LIF	Factor inhibidor de leucemias (siglas en inglés)

MAD	Proteína de dimerización 1 de Max
MAGEB1	Miembro B1 de la familia MAGE
MAX	Factor X asociado a Myc
MCTS	Esferoides Multicelulares Tumoraes (siglas en inglés)
MEIG1	Proteína 1 asociada a la meiosis/espermatogénesis (siglas en inglés)
MEK	Proteína quinasa activada por mitógenos (siglas en inglés)
MEM	Medio Mínimo Esencial Eagle
mESCs	Células embrionarias de ratón (siglas en inglés)
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
MYC	Homólogo del oncogén viral de la mielocitomatosis aviar v-Myc
MYOD	Proteína de diferenciación miogénica (siglas en inglés)
MYT1L	Factor de transcripción de mielina tipo 1 (siglas en inglés)
NANOG	Factor de transcripción homeobox Nanog
NFKB	Factor nuclear potenciador de las cadenas ligeras kappa de las células B activadas
NOTCH	Receptor NOTCH
ns	No significativo
OCT4	Factor de transcripción 4 de enlace octámero
OLF1	Proteína olfativa de tipo receptor 1 (siglas en inglés)
p21	Inhibidor 1A de quinasas dependiente de ciclina
p27	Inhibidor 1B de quinasas dependiente de ciclina
p38 MAPK	Proteína quinasa activada por mitógenos p38 (siglas en inglés)
PAGE	Gel de electroforesis de poliacrilamida (siglas en inglés)
PAS	Dominio Per-Arnt-Sim
PBS	Tampón fosfato salino (siglas en inglés)
PCNA	Antígeno nuclear de proliferación celular (siglas en inglés)
PCR	Reacción en cadena de la polimerasa (siglas en inglés)
PE	Endodermo primitivo (siglas en inglés)
PEG3	Gen paternalmente expresado 3 (siglas en inglés)
PER	Proteína reguladora del ciclo circadiano
PFA	Paraformaldehído (siglas en inglés)
PGCs	Células primordiales germinales (siglas en inglés)
PI	Yoduro de propidio (siglas en inglés)
PI3K	Fosfoinositol 3 quinasa (siglas en inglés)
POU5F1	Factor de transcripción 1 dominio POU clase 5
PPAR γ	Receptor de peroxisoma-proliferador-activado gamma (siglas en inglés)
PTX1	Factor de transcripción de homeodominio tipo pareado
qPCR	PCR cuantitativa (siglas en inglés)
RNA-Seq	Secuenciación de ARN (siglas en inglés)
RT-PCR	Retrotranscripción

RXRB	Receptor de retinoides X Beta (siglas en inglés)
SCR	Vector control de shRNA
SDS	Dodecil Sulfato Sódico
SEM	Error estándar de la media (siglas en inglés)
sgRNA	Guía pequeña de ARN para CRISPR
shARN	Pequeño ARN de horquilla
siARN	Pequeño ARN de interferencia
SIM	Proteína single-minded (siglas en inglés)
SMAD	Proteína homóloga al gen "madres contra decaentaplégico" (siglas en inglés)
SOHLH2	Basic Helix Loop Helix 2 específico de espermatogénesis y oogenénesis
SOX2	Región determinante del sexo Y-box 2
SRA	Archivo de lecturas secuenciadas (siglas en inglés)
STAT3	Transductor de señales y activador de la transcripción 3
STELLA	Proteína asociada a pluripotencia y desarrollo
SWI/SNF	SWIch/Sacarosa no fermentable (siglas en inglés)
SYK	Tirosín-quinasa asociada a bazo
TAE	Solución tamponada con Tris, acetato y EDTA
TAL1	Proteína de leucemia linfoblástica aguda 1 de células T (siglas en inglés)
T-ALL	Leucemia linfoblástica aguda de células T (siglas en inglés)
TBS	Solución salina tamponada con Tris-HCl
TBX3	Factor de transcripción T-Box 3 (siglas en inglés)
TCF4	Factor de transcripción 4 (siglas en inglés)
TCFL5	Factor de transcripción tipo 5 (siglas en inglés)
TE	Trofoectodermo
TGFβ	Factor de crecimiento transformante beta 1 (siglas en inglés)
TLR	Receptor tipo Toll (siglas en inglés)
TP53	Proteína tumoral p53
TWEAK	Débil inductor de la apoptosis relacionada con TNF (siglas en inglés)
Twist1	Factor de transcripción bHLH de la familia Twist 1
USF1	Factor de transcripción upstream 1 (siglas en inglés)
UTR	Región no traducida de los ARNm (siglas en inglés)
VIH	Virus de la inmunodeficiencia humana
WB	Western-blot
WNT	Familia de sitio de integración MMTV tipo-wingless
WT	Wild Type
X-Gal	5-bromo-4-cloro-3-indolil-β-D-galactopiranosido
YBX1	Proteína de unión a regiones Y-box 1
Zfp238	Proteína dedos de Zinc 238 (siglas en inglés)

Resumen / Abstract

Resumen

Los factores de transcripción son proteínas clave en el control de la expresión génica y tienen un papel importante en diversos procesos celulares. La desregulación de estas proteínas está asociado a numerosos procesos patológicos entre los que se encuentra el cáncer. El cáncer es la principal causa de muerte en países desarrollados, siendo el de colon el segundo en incidencia y mortalidad. Este tipo de cáncer es resultado de la alta proliferación de las células epiteliales de la mucosa del colon y sus glándulas, considerándose uno de los más peligrosos por su alta capacidad metastásica y la adquisición de quimioresistencia. *Transcription Factor Like 5* (TCFL5) es un factor de transcripción poco estudiado implicado principalmente en espermatogénesis, sistema inmune y cáncer. Se ha descrito la presencia de dos isoformas: TCFL5 y CHA. Sin embargo, pocos estudios discriminan entre estas. En este trabajo hemos descrito que *TCFL5* es un gen con múltiples isoformas tanto en humanos como en ratón. Nuestros resultados muestran la importancia de la expresión de estas isoformas en cáncer de colon, ya que la delección completa de las isoformas reduce drásticamente las propiedades tumorales de las células, principalmente de la formación de MCTS y del crecimiento tumoral *in vivo*. TCFL5 y CHA presentaron una regulación recíproca y su sobreexpresión específica mostró funciones diferentes en la célula ejerciendo CHA un papel más protumoral que TCFL5. Además, la expresión TCFL5/CHA es fundamental para la actividad NFkB2 regulando la expresión de genes antiapoptóticos como *BCL2*. TCFL5/CHA controla también la expresión de los marcadores de pluripotencia SOX2, NANOG y KLF4. Por último, TCFL5 se encuentra expresado en el desarrollo embrionario temprano de ratón durante el periodo de preimplantación y juega un papel en la diferenciación de células embrionarias a precursores de la línea germinal reprimiendo la expresión de genes de pluripotencia y controlando la de genes importantes en la diferenciación. En conjunto, este trabajo pone de manifiesto la importancia de TCFL5 en el desarrollo de cáncer de colon y procesos de pluripotencia y diferenciación. Además, plantea un complejo modelo de regulación de su expresión donde el ratio de expresión determina el fenotipo final de la célula.

Abstract

Transcription factors are key proteins in the gene expression control. They have an important role in several cellular processes and their dysregulation is associated to many pathological processes such as cancer. Cancer is the first cause of death in developed countries and specifically colorectal cancer has the second highest incidence and mortality. This cancer is a consequence of increased proliferation of epithelium and its glands. Because of its metastatic and chemoresistance capacity it is one of the most dangerous cancer. Transcription Factor Like 5 (TCFL5) is a poorly studied transcription factor implicated in spermatogenesis, immune system and cancer. Two isoforms for this gene have been found: TCFL5 and CHA. However, few studies discriminated them. In this work we have described multiple isoforms for *TCFL5* in humans and mice. Our results showed the relevance of these isoforms expression in colorectal cancer. Complete deletion of the isoforms drastically decreased cell tumoral properties, mainly the *in vivo* tumoral growth and multicellular tumor spheroids (MCTS) formation. TCFL5 and CHA showed a reciprocal regulation and its specific over-expression presented different functions: CHA had more pro-tumoral effects than TCFL5. Moreover, TCFL5/CHA over-expression was important for the NFkB2 activity regulating anti-apoptotic genes as *BCL2*. These isoforms also controlled pluripotency markers expression such as SOX2, NANOG and KLF4. Finally, TCFL5 was expressed in mouse early embryonic development during preimplantation stages. It also had a role in the embryonic stem cells differentiation to primordial germinal cells repressing the pluripotency genes expression and controlling differentiation genes. Collectively, TCFL5 is an important gene in the colorectal cancer development, pluripotency and differentiation processes. In addition, this work shows a complex regulation model where the ratio of expression determines the cellular phenotype.

Introducción

Introducción

1. Factores de transcripción.

El dogma de la biología molecular postulado por Francis Crick en 1957 defiende que la información genética se transfiere de manera secuencial de ácidos nucleicos a proteína [1]. Aunque los nuevos avances han obligado a actualizar este postulado y se ha aumentado el número de componentes de este, este dogma consta de tres procesos principales: replicación, transcripción y traducción.

El fenómeno de la transcripción es un proceso complejo mediante el cual la información codificada en el genoma se convierte en ARN. En eucariotas, dependiendo de la complejidad del organismo, desde un 5% hasta un 10% del ADN codificante contiene la información necesaria para producir proteínas que regulan la transcripción. Estas proteínas se dividen en: proteínas de unión al ADN que median la activación o represión selectiva de genes; proteínas que pertenecen a grandes complejos de la maquinaria de la ARN-polimerasa para iniciar la transcripción; y proteínas modificadoras y reguladoras de la cromatina [2]. El primer grupo hace referencia a los factores de transcripción que se definen como proteínas con capacidad de unirse a secuencias específicas del ADN y regular la transcripción [3]. Aunque para la mayoría de factores de transcripción continúa hoy sin caracterizarse cuál es su función, en general tienen un papel clave en el control del desarrollo, procesos celulares, transducción de señales, respuesta a estímulos, procesos reproductivos y sistema inmune [4].

1.1. Clasificación de los factores de transcripción.

Los factores de transcripción se clasifican según su dominio de unión al ADN. Desde su descubrimiento, se han clasificado en 4 grandes grupos: *helix-turn-helix*, dedos de zinc, cremalleras de leucinas y motivos de unión sin clasificar [5] [6]. A medida que han avanzado las técnicas, a lo largo de los años se han ido determinando en alta calidad la estructura de las proteínas de unión a ADN y estos grupos o superclases se han ido actualizando. Actualmente existen 8 superclases estructurales/funcionales basados en dicha clasificación los cuales han sido clasificados en 54 familias estructurales [7].

El grupo I está representado por las proteínas *helix-turn-helix*. El dominio de estas proteínas se caracteriza por 2 α -hélices perpendiculares conectadas por un giro- β o un loop de longitud suficiente para que las α -hélices queden enfrentadas [8]. Al grupo II pertenecen las proteínas coordinadoras de zinc cuyo motivo de unión al ADN se caracteriza por una coordinación tetrahédrica de uno o dos iones de zinc mantenidos por residuos de cisteína e histidina. Este dominio no es exclusivo de proteínas de unión al ADN. El grupo III lo representan proteínas tipo cremallera exclusivas de organismos eucariotas. Su nombre proviene de la forma de dimerización de sus miembros y está compuesto entre

otros por las clases de proteínas cremallera de leucina y *basic helix-loop-helix* (bHLH). El grupo IV lo componen otras proteínas α -helix que, aunque todas utilizan una α -helix para unirse al ADN, su estructura es muy variada. El grupo V está formado por proteínas lámina- β y únicamente compuesta por proteínas de unión a la región TATA-box. El grupo VI son proteínas β -horquilla. El grupo VII lo componen otras proteínas las cuales no han sido bien determinadas su estructura y no presentan actividad enzimática. Por último, el grupo VIII lo forman enzimas que más que tener una estructura en común, presentan actividad enzimática que altera la estructura del ADN [7].

1.2. Basic Helix Loop Helix.

Dentro de la superclase III de factores de transcripción se encuentran los bHLH. Estos factores de transcripción se caracterizan por la presencia de un motivo básico de unión al ADN, encargado de reconocer la secuencia E-box canónica CANNTG, y un motivo HLH por el cual pueden dimerizar formando homodímeros o heterodímeros [9]. La clasificación de estos factores de transcripción ha ido variando según se avanzaba en su conocimiento pasando de 4 familias [10] a 6 familias atendiendo a su evolución, las secuencias E-box a las que se unen, residuos conservados en otras partes del dominio y la presencia o ausencia de motivos adicionales [11].

La primera familia (familia A) se compone de los bHLHs que reconocen las secuencias E-box CACCTG o CAGCTG, compuesto por proteínas como MYOD y TWIST, mientras que la familia B reconocen las secuencias CACGTG o CATGTTG, compuesto por proteínas como MYC, MAX y MAD. La tercera familia o C presenta un dominio PAS y reconoce la secuencia ACGTG o GCGTG. Este grupo puede dimerizar con otras proteínas con o sin dominio PAS y con pequeñas moléculas. A esta familia pertenecen proteínas como ARNT y HIF. La familia D, lo componen las proteínas ID que carecen de la región básica por lo que no se unen al ADN y actúan como inhibidores tras la dimerización. La familia E lo componen los bHLH-O que se une preferentemente a regiones N-box del tipo CACGCG o CACGAG, contienen un dominio orange y generalmente presentan una función represora. A este grupo pertenecen los factores de transcripción HEY1 y HEY2. Por último, a la familia F pertenecen las proteínas COE que se caracterizan por la presencia de un dominio COE involucrado tanto en la dimerización como en la unión al ADN [12].

A lo largo de la evolución los bHLH aparecen sólo en eucariotas, de levaduras a humanos, y su número aumenta con la complejidad del organismo lo que refleja su papel en el desarrollo de la multicelularidad, lo que implica su función en procesos metabólicos y de desarrollo requeridos para dicha condición [11]. La diversidad de secuencias de unión al ADN y la capacidad de dimerizar proporcionan a estos factores de transcripción una enorme plasticidad y diversidad de genes que regular. Existen bHLH expresados ubicuamente en todos los tejidos asociados a la regulación de procesos celulares como proliferación y apoptosis mientras que algunos bHLH son expresados de manera específica de tejido regulando procesos de determinación del destino celular y diferenciación

como miogénesis, neurogénesis o hematopoyesis [13] [12].

1.3. Control de la actividad de los factores de transcripción.

Múltiples factores de transcripción se pueden unir formando dímeros o complejos proteicos de manera cooperativa a sus secuencias diana en el ADN en regiones denominadas *enhancer* regulando la actividad de promotores cercanos y por tanto la transcripción de su gen diana. Además, los factores de transcripción también pueden regular la transcripción de genes más alejados a través del contacto físico involucrando a proteínas que generan pliegues en la estructura del ADN acercando el enhancer y el promotor [14]. La región promotora que presenta las secuencias necesarias para que la transcripción ocurra y que incluye al complejo ARN-polimerasa II también puede presentar unión de ciertos factores de transcripción [15].

El control de la actividad de los factores de transcripción es clave para el correcto funcionamiento celular. Estas proteínas son los últimos efectores de las cascadas de señalización y son las encargadas de activar o reprimir la expresión de genes específicos. La regulación de la actividad de estas proteínas puede darse a diferentes niveles, desde la propia transcripción del factor de transcripción, transporte del ARNm y traducción, hasta modificaciones post-traduccionales y unión al ADN (Figura 1).

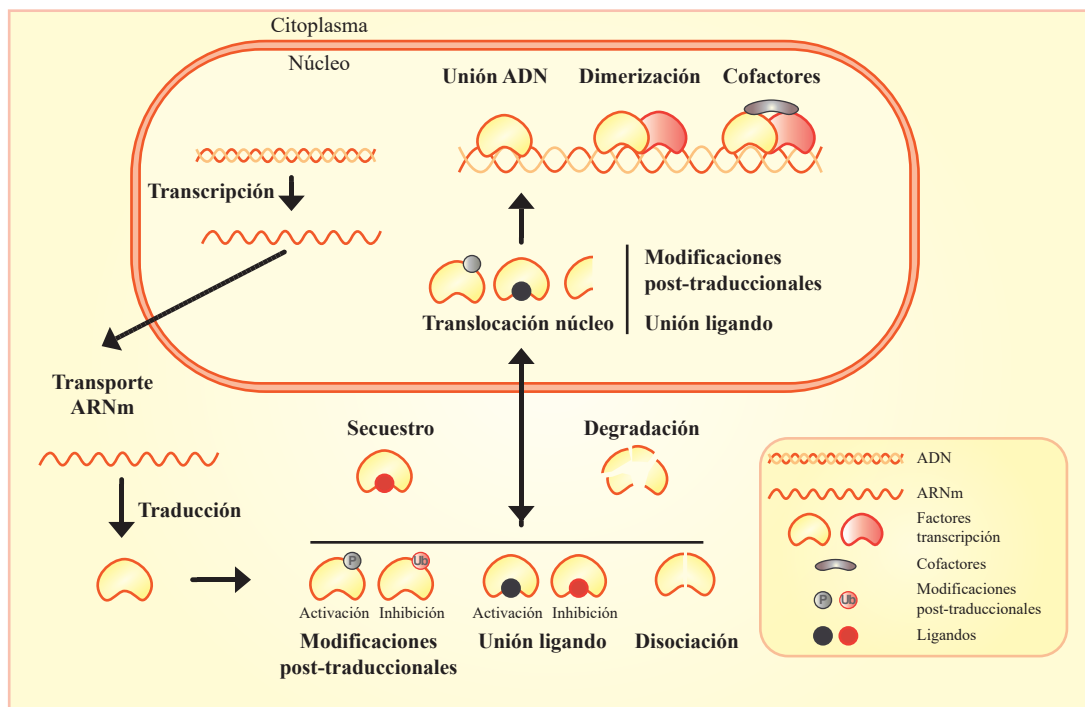


Figura 1. Esquema del control de la actividad de los factores de transcripción. Se representan los distintos puntos en los cuales un factor de transcripción puede regular su actividad, desde su expresión, transcripción, transporte al citoplasma y traducción, hasta activación por modificaciones traduccionales, entrada al núcleo y formación de complejos transcripcionales. Esquema adaptado de *Calkhoven and Ab 1996*.

La activación o represión de un gen puede darse con la simple unión del factor de transcripción a su región del ADN. Sin embargo, en la mayoría de ocasiones el escenario es más complejo. De manera indirecta un factor de transcripción puede competir por el lugar de unión de otro factor de transcripción y bloquear así su función. También puede ser secuestrado o inactivado por la formación de dímeros o complejos proteicos [16]. En la mayoría de los casos, la actividad de los factores de transcripción depende de modificaciones post-traduccionales por lo que la expresión no indica necesariamente una actividad de los mismos [4]. Estas modificaciones, que incluyen fosforilaciones, acetilaciones, metilaciones, sumoilaciones o ubiquitinaciones entre otras, pueden tener diferentes efectos como la degradación, la entrada o salida del núcleo, o la disociación de un complejo o de una región inhibidora [17] [18]. Por otra parte, aunque la unión de los factores de transcripción al ADN es necesaria para el inicio de la transcripción, el proceso requiere la unión de cofactores que junto a los factores de transcripción reclutaran a la ARN-polimerasa [19].

Con respecto a los bHLH, la dimerización de estas proteínas representa uno de los puntos de control de su actividad. Estos factores de transcripción pueden formar homodímeros o heterodímeros entre miembros de la misma o distinta familia. La formación de estos dímeros aumenta el número de regiones de ADN al que pueden unirse incrementando los genes sobre los que ejercen su función y por tanto su efecto final [20]. Un ejemplo de la importancia de la regulación de función por la formación de dímeros son los complejos MYC-MAX y MAD-MAX que afectan a diferentes procesos celulares como proliferación, diferenciación y apoptosis presentando en muchas ocasiones efectos antagónicos. MAX es un factor de transcripción ubicuo mientras que la expresión de MYC y MAD responde a diversas señales de señalización. El dímero MYC-MAX recluta al promotor de sus genes diana al complejo SWI/SNF, un complejo remodelador de cromatina, y a Histonas Acetil Transferasas (HATs) provocando una apertura de la cromatina y favoreciendo el inicio de la transcripción. Por su parte, el dímero MAD-MAX recluta a histonas deacetilasas (HDACs) produciendo el efecto contrario [21] [22]. Además, la formación de dímeros también puede regularse a su vez por modificaciones post-traduccionales. Es el caso del dímero E47-MYOD necesario para el desarrollo del músculo esquelético. La fosforilación del bHLH E47 por parte de p38 MAPK bloquea la formación del homodímero E47 favoreciendo la formación de heterodímero E47/MYOD [23].

2. Cáncer

El cáncer es la principal causa de muerte en países desarrollados. Se estima que, en 2018, aproximadamente 18 millones de nuevos casos de cáncer fueron diagnosticados en todo el mundo y cerca de 10 millones de personas murieron por esta enfermedad [24]. Sólo en España, cerca de 228.000 nuevos casos fueron detectados en 2017 y se estima que para 2035 el número de nuevos diagnósticos ascenderá a 315.000 [25]. Mundialmente, el cáncer de pulmón es el más diagnosticado y la principal causa de muerte por este tipo de enfermedad tanto en hombres como en mujeres. En segundo lugar se encuentra el cáncer de colon tanto en diagnóstico como en mortalidad en ambos

sexos [24].

Según el *National Cancer Institute* (NIH), el cáncer es el nombre dado a una colección de enfermedades relacionadas en las cuales, células anormales se dividen sin control pudiendo llegar a invadir tejidos cercanos [26]. El desarrollo del cáncer se basa en dos procesos: la continua adquisición de variaciones genéticas heredables en células individuales debido a mutaciones más o menos aleatorias, y a una selección natural que actúa sobre el fenotipo resultante. Generalmente, esas alteraciones pasan desapercibidas o se ponen de manifiesto como crecimientos benignos. Sin embargo, puntualmente, una célula adquiere suficientes alteraciones ventajosas como para proliferar de manera autónoma, invadir tejidos y metastatizar [27].

Durante décadas, los diversos trabajos sobre esta enfermedad han revelado un escenario con una biología muy compleja que abarca la desregulación de multitud de procesos y que se han reunido bajo lo que se conoce como las características del cáncer o *hallmarks of cancer*. Según esta clasificación, el desarrollo del cáncer implica una señal continua de proliferación, evasión de la supresión del crecimiento, evasión de la respuesta inmune, inmortalidad replicativa, inflamación, invasión y metástasis, inducción de la angiogénesis, inestabilidad genómica y mutaciones, resistencia a la muerte celular y desregulación de la energía celular [28].

2.1. Factores de transcripción en cáncer

El cáncer se produce por alteraciones genéticas que, consecuentemente, llevan a una desregulación del programa de transcripción. La alteración del programa transcripcional puede darse a varios niveles, desde modificaciones de la cromatina que vuelven a una región promotora activa o inactiva alterando así su transcripción, hasta alteraciones de los factores de transcripción. Los factores de transcripción son piezas clave de la expresión génica y, por tanto, su desregulación tiene consecuencias fisiológicas que repercuten en el desarrollo de diversas patologías como el cáncer [29] (Figura 2).

Existen factores de transcripción claves, denominados pioneros o *master transcription factors*, controlando la expresión de la red transcripcional que define la célula y cuya alteración modifica dicha red [30]. Es el caso del factor de transcripción bHLH TAL1. TAL1 se expresa y se silencia de manera específica durante el desarrollo temprano de timocitos. Entre el 40% y el 60% de las leucemias linfoblásticas agudas de células T (T-ALL) presentan una expresión elevada de esta proteína. Su sobreexpresión induce un núcleo de genes expresados normalmente en el desarrollo temprano de estas células, reprograma a estas células a un estado menos diferenciado y contribuye en gran medida al desarrollo y mantenimiento de la leucemia T-ALL [31].

Por otro lado, existen factores de transcripción que no alteran la red transcripcional si no que

amplifican o silencian la ya existente. Un ejemplo es el factor de transcripción miembro de los bHLH cMYC. cMYC es un oncogén que se encuentra frecuentemente elevado en muchos tipos de cáncer y se asocia al desarrollo de la enfermedad, aunque no es capaz de inducirla por sí solo. Se ha observado que altos niveles de cMYC llevan a un aumento de la unión a los promotores de sus genes diana, aumentando su transcripción, más que activar la expresión de nuevos genes [32] [33]. Por tanto, cMYC no se considera un alterador de la red transcripcional, sino que amplifica la red existente. Las funciones atribuidas a cMYC son muy diversas y abarcan desde proliferación y apoptosis hasta síntesis de proteínas, metabolismo celular y evasión de la respuesta inmune explicando que su desregulación tenga efectos pleiotrópicos en el cáncer [34].

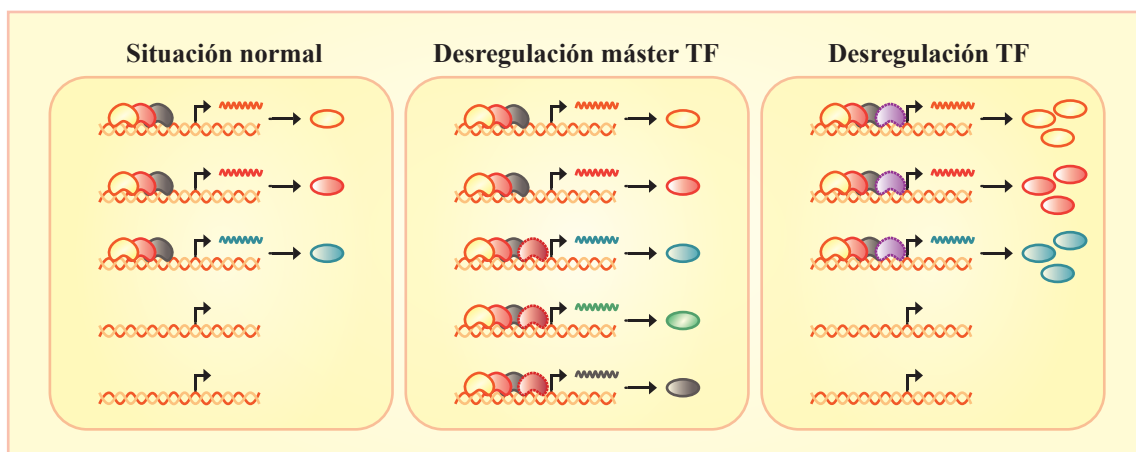


Figura 2. Esquema de la desregulación de los factores de transcripción en cáncer. A la izquierda, la situación de expresión normal en un tejido sano. En el centro, una desregulación de un máster TF modificando la red transcripcional de la célula. A la derecha, la desregulación de un factor de transcripción que amplifica la red transcripcional de la célula. Esquema adaptado de Bradner, Hnisz et al. 2017.

2.2. Desregulación de la actividad de los factores de transcripción en cáncer.

Como se ha mencionado anteriormente, la actividad de los factores de transcripción se regula a distintos niveles bien sea a nivel de su propia expresión, modificaciones post-traduccionales o unión al ADN. Estas alteraciones en la actividad se pueden deber a amplificaciones, deleciones o reordenamientos cromosómicos del ADN, modificando su expresión, o a mutaciones puntuales que conllevan una pérdida o ganancia de función afectando a su actividad o unión al ADN [35]. Por tanto, la función de estas proteínas se puede ver alterada de manera diversa y afectar a cualquiera de estos puntos.

Un ejemplo de alteración en la expresión de un factor de transcripción es cMYC en el linfoma de Burkitt. En este tipo de leucemia de células B, cMYC es translocado cerca de la región *enhancer* de la cadena pesada de las inmunoglobulinas [36] [37]. De esta forma, la expresión de esta proteína no se encuentra bajo el control de su promotor sino bajo el promotor de la cadena pesada de las inmunoglobulinas que se expresa continuamente en estas células. Como consecuencia, se produce una expresión constante de cMYC produciendo un aumento de sus niveles de expresión y

desregulando la transcripción de sus genes diana [38].

En cuanto a alteraciones en la actividad debido a mutaciones puntuales, el gen supresor de tumores *TP53*, codificante para el factor de transcripción p53, es un ejemplo. La activación de p53 en respuesta a diversos estímulos generalmente contribuye a la supresión del crecimiento tumoral. Sin embargo, p53 se encuentra mutado en un gran número de tumores y estas mutaciones, en su mayoría, generan la pérdida de su función normal favoreciendo la formación del tumor, invasión y metástasis. Sin embargo, se ha observado que estas mutaciones también pueden generar ganancia de función [39] [40]. Además, ciertas mutaciones de p53 pueden actuar como dominante negativo del p53 *wild type*. Mediante la transfección de un plásmido que expresa en cantidades equitativas la forma *wild type* de p53 y distintas mutaciones de p53, se observó que, en algunas funciones como apoptosis, ciertas mutaciones actúan como dominante negativo inhibiendo la transcripción de genes de la familia BAX mientras que en otras como proliferación no tienen efecto [41]. Este ejemplo es una muestra del amplio abanico de posibilidades que las mutaciones pueden tener sobre la actividad de un factor de transcripción atendiendo al tipo y lugar de estas mutaciones.

2.3. Los factores de transcripción como diana terapéutica contra el cáncer.

El papel que desempeñan los factores de transcripción en el control de la expresión génica hace de estas proteínas unas dianas interesantes para el tratamiento contra el cáncer. El concepto de terapia transcripcional se acuñó como una estrategia para revertir el efecto de una transcripción aberrante en células cancerígenas mediante la intervención directa en el proceso de transcripción. Comparando con los tratamientos clásicos que afectan a vías de señalización, la transcripción se consideraba un proceso no tratable debido a la localización nuclear del proceso y por tanto de difícil acceso para los agentes terapéuticos, y lo esencial de sus componentes que pueden afectar a diversos procesos y generar efectos secundarios no deseados [42]. Sin embargo, debido al avance en los conocimientos de los factores de transcripción implicados en procesos carcinogénicos tales como su función en cáncer, expresión, degradación, interacciones proteína-proteína e interacciones proteína-ADN, se ha logrado el desarrollo de nuevas estrategias que hacen de los factores de transcripción agentes terapéuticos contra el cáncer [43].

Diferentes estrategias se han utilizado contra los factores de transcripción para su inhibición o activación atendiendo a diferentes aspectos. Una de estas estrategias ha sido la inhibición de la expresión de los factores de transcripción a través de compuesto que afecten a la transcripción directa de estas proteínas, alterando su estado epigenético o bloqueando su maquinaria de transcripción. También se utilizan compuesto que llevan a los factores de transcripción a degradación vía proteasoma a través de modificaciones post-transcripcionales. Además, se han utilizado estrategias que alteran las interacciones proteína-proteína y proteína-ADN de los factores de transcripción. Las interacciones proteína-proteína se alteran a través de compuestos que impiden la formación, bien de dímeros o

bien de complejos proteicos, a través del secuestro de la proteína, estabilización o degradación de estas interacciones. Con respecto a la interacción proteína-ADN, la mayoría de compuestos alteran la estructura del ADN y como consecuencia secundaria impiden la transcripción. Sin embargo, cada vez son más los compuestos de este tipo específicos para secuencias determinadas que alteran la unión de un determinado factor de transcripción. Por último, se han diseñado compuestos que afectan a la región catalítica de los factores de transcripción impidiendo su actividad [43].

3. Cáncer de colon.

Según la Sociedad Americana contra el Cáncer, en 2017, sólo en Estados Unidos alrededor de 130.000 personas fueron nuevamente diagnosticadas con cáncer de colon y unas 50.000 murieron por esta enfermedad [44]. En España, un informe elaborado en 2018 por la Asociación Española Contra el Cáncer (AECC) muestra que el cáncer colorrectal es el primer cáncer diagnosticado en España con cerca de 35.000 nuevos casos detectados en 2017, siendo el segundo en mortalidad tanto en mujeres como hombres con alrededor de 16.000 defunciones al año [45]. Los principales factores de riesgo para el cáncer colorrectal son la edad y el componente genético. Por otro lado, aspectos modificables como la inactividad física, el tabaquismo, sobrepeso o consumo excesivo de carne roja son considerados también factores de riesgo [46].

La cirugía, quimioterapia y radiación son las principales estrategias terapéuticas contra el cáncer de colon. No obstante, la cirugía no es viable en casos avanzados por lo que la quimioterapia sigue siendo la opción más utilizada. Sin embargo, en un gran porcentaje de casos el tratamiento no es efectivo, aunque ha conseguido prolongar la esperanza de vida en torno a 40 meses desde el diagnóstico. Actualmente, nuevas estrategias terapéuticas como la inmunoterapia están siendo utilizadas en ensayos clínicos con la esperanza de llegar a buen término [47].

3.1. Tipos de cáncer de colon.

El cáncer colorrectal es un tumor maligno, o neoplasia, resultado de la proliferación excesiva de las células epiteliales de la mucosa del colon y de sus glándulas junto de una inhibición progresiva de la muerte celular [48]. En función de la causa que lo origina se puede dividir al cáncer colorrectal en dos tipos: hereditario o familiar y esporádico.

Aproximadamente el 30% de los pacientes con cáncer colorrectal presenta antecedentes familiares, aunque sólo un 5% tienen un componente hereditario realmente identificable [49]. El cáncer colorrectal hereditario se divide en dos tipos basados en la presencia de pólipos: poliposis adenomatosa familiar, o cáncer colorrectal no polipósido hereditario [50].

El cáncer de colon esporádico representa aproximadamente el otro 70% de los casos

diagnosticados. Se consideran dos mecanismos principales por los cuales se desarrolla este tipo de cáncer. El primer mecanismo, denominado canónico o secuencia adenoma-carcinoma, representa entre el 80% y el 85% de los casos esporádicos. Este mecanismo implica inestabilidad cromosómica y lleva a alteraciones de los genes supresores de tumores *APC*, *TP53* y *DDC* y del oncogén *KRAS*. El segundo mecanismo denominado mutador representa entre el 15% y el 20% de los casos esporádicos e involucra a la inestabilidad de los microsatélites [51].

Para el mecanismo canónico inicialmente se propuso un modelo de progresión en el cual la mutación de al menos 4 de los genes citados es condición necesaria para el inicio del proceso tumoral maligno. Además, se propuso que aunque la alteración de estos genes suele ocurrir preferentemente en un determinado orden, es la acumulación de estas y no el orden las responsables de las propiedades biológicas del tumor [52]. Posteriormente, este modelo ha sido modificado indicando que las alteraciones genéticas de dichos genes se suceden en un orden estricto. Por otro lado, se han involucrado a las vías de señalización de Wnt y TGF β como parte crucial de este proceso [53] (Figura 3).

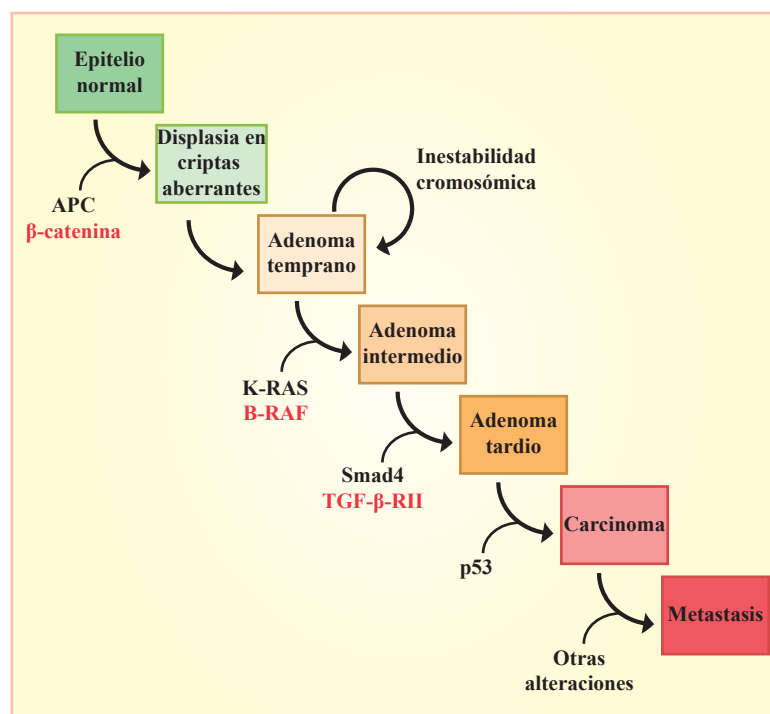


Figura 3. Esquema resumen que combina las alteraciones genéticas con el desarrollo del cáncer colorrectal. Cada fase del desarrollo tumoral está promovida por la mutación de uno o varios genes. Esquema adaptado de *Fearon and Vogelstein, 1990*.

El mecanismo mutador se caracteriza por una alta acumulación de mutaciones en las secuencias microsatélites. Estas mutaciones son causadas por una alteración en los mecanismos de reparación de daño del ADN de la célula. Comparado con el mecanismo canónico, este no implica grandes anomalías citogenéticas descritas en los principales genes propuestos en el modelo canónico. Sin embargo, se han descrito mutaciones en numerosos genes implicados en la progresión del cáncer colorrectal como *APC*, *TCF4* o *CDX2* [51].

3.2. Factores de transcripción y cáncer de colon.

Diversos estudios muestran un cambio en la expresión génica durante la progresión del cáncer colorrectal, desde lesiones precancerosas hasta carcinomas, encontrándose afectados distintos procesos celulares como la progresión del ciclo celular, homeostasis y diferenciación celular [54] [55]. Esto muestra un cambio en la red transcripcional durante el desarrollo del proceso tumoral. En consonancia con esto, en un estudio de muestras de pacientes de adenocarcinoma frente a tejido sano, se ha observado que la red transcripcional en adenocarcinomas presenta una alteración en la expresión de al menos 250 factores de transcripción indicando una fuerte desregulación de la transcripción durante este proceso tumoral [56]. Además, se ha asociado una sobreexpresión de determinados factores de transcripción al proceso de metástasis de este tipo de cáncer [57].

4. Desarrollo embrionario temprano.

El periodo de preimplantación del desarrollo temprano embrionario comprende desde la fecundación del óvulo, que resulta en la formación del cigoto, hasta la blástula, capaz de implantarse en el útero donde continuará el desarrollo embrionario. Durante este proceso se producen una serie de divisiones celulares dentro de la zona pelúcida, sin aumentar el tamaño total del embrión [58]. Desde el cigoto, estadio de 1 célula, se suceden los estadios de 2 células, 4 células y 8 células. A partir del estadio de 8 células se produce la polarización y la compactación seguida de divisiones asimétricas para dar lugar al trofoectodermo (TE), que contribuirá a la placenta, y la masa celular interna (ICM) en el estadio de mórula. Ya en el estadio de blástula, la ICM dará lugar al epiblasto (EPI), que en etapas más tardías producirá los linajes del ectodermo, endodermo y mesodermo y por tanto todas

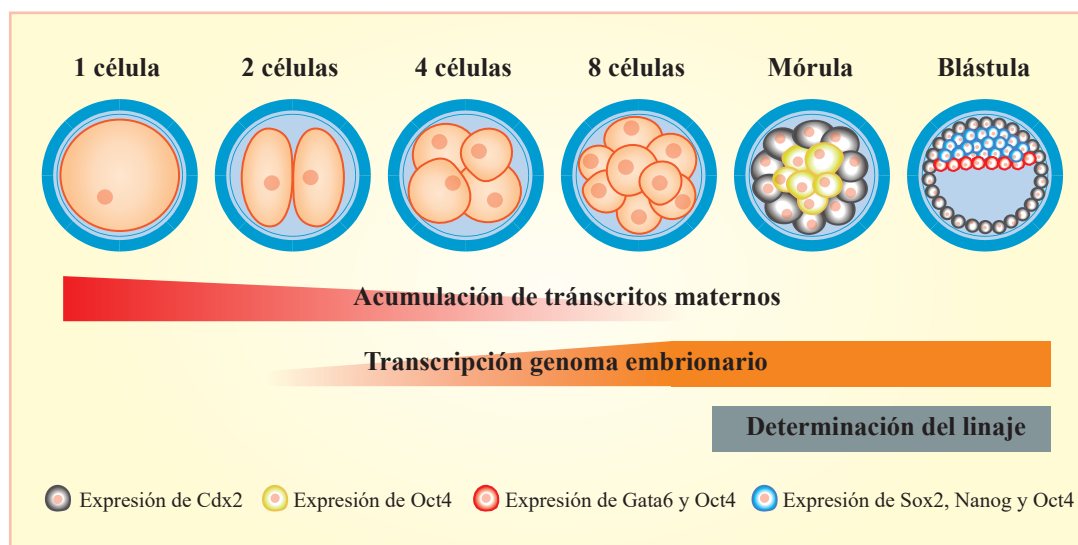


Figura 4. Estadios del periodo de preimplantación de ratón. Se representan los estadios de 1 célula, 2 células, 4 células, 8 células y blástula. Los transcritos maternos se van degradando a lo largo del proceso mientras que tiene lugar la activación de la transcripción del genoma del embrión. A partir de mórula comienza la determinación de linaje celular: trofoectodermo (célula gris) y masa celular interna (célula amarilla). Ya en blástula, la masa celular interna dará lugar al epiblasto (célula azul) y al endodermo primitivo (célula roja). Esquema adaptado de Mihajlovic and Bruce 2017.

las células fetales, y al endodermo primitivo (PE), que contribuirá mayoritariamente al saco vitelino extra-embionario. Cada célula del embrión se denomina blastómero y hasta el estadio de 8 células se ha demostrado que mantienen la capacidad de totipotencia, es decir, dar lugar a los 3 linajes celulares primarios [59]. La formación de las uniones celulares, la compactación y la polarización de la célula a partir de la mórula, son procesos indispensables y definen la primera decisión del linaje celular [60] (Figura 4).

4.1. Factores de transcripción en etapas tempranas del desarrollo embrionario.

Durante el periodo de preimplantación, mientras que el ARNm materno se va degradando, se produce una activación de los genes del embrión en un proceso que se denomina activación del genoma del cigoto. Esta activación tiene lugar en dos olas, una inicial menor en las primeras etapas y una posterior denominada mayor. Este fenómeno es crucial para expresar las proteínas necesarias para comenzar los distintos procesos celulares que se dan durante el desarrollo [61]. En este punto, los factores de transcripción juegan un papel clave en los cambios de expresión observados durante este periodo. Estudios recientes muestran un cambio significativo en la expresión de los factores de transcripción en el estadio de 4 células a 8 células coincidiendo con este fenómeno de activación del genoma del cigoto. En los primeros estadios, solo unos pocos factores de transcripción se encuentran expresados mientras que, en etapas más tardías, el número de factores de transcripción va aumentando significativamente a medida que avanza el desarrollo. Además, la dinámica de expresión de algunos factores de transcripción oscila a lo largo del proceso indicando un papel en determinado momento [62].

Aunque similar, el desarrollo embrionario entre mamíferos no es exactamente igual. Existen diferencias específicas de especies. Entre humanos y ratón se han observado diferencias en el tiempo de la ola mayor de la activación del genoma del cigoto, en patrones de expresión y en patrones epigenéticos [63]. Aunque las redes transcripcionales parecen mantenerse conservadas, las vías de señalización que regulan la diferenciación de linaje pueden variar. Además, algunos estudios sugieren el papel de los factores de transcripción en las diferencias entre especies indicando variaciones en las secuencias de unión de los factores de transcripción en los promotores [64] [65].

4.2. Los factores de transcripción OCT4, NANOG y SOX2 durante el desarrollo embrionario temprano.

Los factores de transcripción son claves para el correcto control genético durante el desarrollo y son muchos los implicados en el periodo de preimplantación. Sin embargo, algunos de los más estudiados por su papel clave en este proceso y en pluripotencia son OCT4, NANOG y SOX2.

OCT4, también denominado POU5F1, pertenece a la superclase *helix-turn-helix* y aunque

su expresión es uniforme en los blastómeros de estadios tempranos, su distribución en estos no lo es. Se ha estudiado la movilidad celular al núcleo de este factor de transcripción durante el periodo de preimplantación y se ha establecido blastómeros con mayor y menor movilidad debido a la unión de OCT4 al ADN. La mayor o menor movilidad de este factor de transcripción se ha asociado con una división simétrica o asimétrica de los blastómeros, respectivamente, y por tanto, al tipo de linaje al que dará lugar [66] [67]. Estos resultados indican que no sólo la presencia, si no la distribución de los factores de transcripción juega un papel clave durante el desarrollo embrionario. OCT4 es un marcador del linaje ICM y su represión es necesaria en el linaje TE. Otro ejemplo de la importancia de los factores de transcripción durante el desarrollo y la determinación celular es CDX2. CDX2 pertenece al grupo de *helix-turn-helix* y es esencial en la formación de TE reprimiendo genes específicos de otros linajes como son OCT4 [68]. Además, CDX2 se une a OCT4 suprimiendo su actividad transcripcional durante la formación del TE [69].

NANOG es un factor de transcripción perteneciente a la familia *homeobox* de la superclase *helix-turn-helix*. Esta proteína es también un marcador de ICM y, junto con el también factor de transcripción GATA6, determinan la decisión de linaje celular de las células de la ICM hacia EPI o TE. La expresión de estas proteínas en los blastómeros de 8 células es uniforme, sin embargo, a partir de blástula, las células de la ICM deciden entre la expresión de NANOG o GATA6 llevando a la diferenciación a linaje EPI o a linaje PE, respectivamente [70] [71]. Además, se ha observado que NANOG inhibe la expresión de GATA6 a través de la unión a su promotor [72].

SOX2 es un factor de transcripción perteneciente a la familia *High Mobility Group* (HMG) de la superclase de otras α -hélice. La expresión de SOX2 es paralela a la de OCT4, se detecta a partir de mórula y es específica de ICM, del posterior linaje EPI y de células de la línea germinal. Además, la expresión de SOX2 se mantiene en las células que darán lugar al sistema nervioso. La delección en el cigoto de SOX2 es letal para el embrión debido a que no se genera un linaje EPI pluripotente [73] (Figura 4).

5. Pluripotencia.

La pluripotencia se define como la habilidad de una célula de auto-renovarse y de dar lugar a las 3 capas germinales primarias (ectodermo, mesodermo y endodermo) del desarrollo embrionario y por tanto a todas las células de un individuo adulto, pero no a los tejidos extra-embrionarios como la placenta [74]. Las células embrionarias de ratón (mESCs) derivan de la ICM formada durante el periodo de preimplantación. Estas células pueden mantenerse *in vitro* indefinidamente conservando su capacidad de pluripotencia bajo unas condiciones de cultivo concretas. Sin embargo, también bajo condiciones específicas, son capaces de diferenciarse a los diferentes linajes celulares o incluso participar de nuevo en el desarrollo del embrión si son de nuevo reintroducidas en este. Estas propiedades se convierten en una herramienta útil no solo para el estudio de la pluripotencia y

diferenciación sino para su uso en medicina [75].

5.1. Factores de transcripción en pluripotencia.

Tanto en humanos como en ratón se ha definido un núcleo central de factores de transcripción que controlan, y por tanto definen, el estado de pluripotencia. Esos factores de transcripción son OCT4, SOX2 y NANOG, todos ellos, marcadores del ICM del embrión. Estas proteínas, junto con otros factores de transcripción y remodeladores de cromatina, forman complejos proteicos que regulan conjuntamente la pluripotencia de las células a través de la activación o represión de diversos genes entre los que se sitúan un gran número de factores de transcripción [76] [77]. Además, no solo controlan la expresión de otros genes, sino que estas proteínas forman una red transcripcional regulando su propia expresión. De esta forma, se ha observado que el complejo OCT4/SOX2 controla la expresión de NANOG uniéndose a su promotor [78] [79] además de controlar la expresión del propio OCT4 a través de la unión a su propio promotor [80]. Por otro lado, se ha observado que SOX2 no sólo regula directamente la expresión de OCT4, también es necesario para la expresión de otros factores de transcripción que regulan a su vez la expresión de OCT4 [81] indicando un complejo mecanismo de autorregulación directo e indirecto.

Sin embargo, aunque estos factores de transcripción controlen la pluripotencia, pequeñas variaciones en la expresión de estas proteínas llevan a las células a diferenciarse. Por ejemplo, la sobreexpresión de OCT4 dos veces por encima de sus niveles habituales lleva a las células hacia la diferenciación a endodermo y mesodermo mientras que el silenciamiento promueve la pérdida de pluripotencia y la formación del trofoectodermo [82]. Además, se ha observado también que el incremento de al menos dos veces de los niveles de expresión de SOX2 conlleva la diferenciación de las células hacia distintos linajes excepto endodermo [83]. Esto demuestra que el equilibrio de la red transcripcional tejida por estos factores de transcripción es crucial para el estado de pluripotencia y pequeñas variaciones desequilibran el sistema alterando la red transcripcional e iniciando los procesos de diferenciación de los distintos linajes.

La sobreexpresión de este núcleo de factores de transcripción en células embrionarias es suficiente para el mantenimiento de la pluripotencia. Sin embargo, la activación de distintas vías de señalización también lleva a la activación de esta red transcripcional. Así, se ha comprobado que en células mESCs el mantenimiento del estado de pluripotencia de estas células requiere de la citoquina *Leukemia Inhibitory Factor* (LIF). LIF previene la diferenciación y promueve la proliferación de las mESCs vía el receptor GP130. La señalización por este receptor lleva a la activación de la ruta vía JAK-STAT3 y por tanto a la activación del factor de transcripción KLF4 mientras que la activación vía PI(3)K-AKT activa al factor de transcripción TBX3. KLF4 y TBX3 mantienen la expresión de SOX2 y NANOG, respectivamente, además de la de OCT4 [84]. Por otro lado, se ha visto que en cultivos libres de suero, LIF por sí sólo no es capaz de mantener el estado de pluripotencia. La

proteína morfogénica ósea 4 (BMP4) es crucial también para el mantenimiento de la pluripotencia. BMP4 vía SMADs induce la expresión de los factores de transcripción represores ID, miembros de la familia D de los bHLHs, ayudando al mantenimiento de la pluripotencia junto con LIF [85].

La importancia de estos factores de transcripción en pluripotencia es tal que la transfección de OCT4, SOX2, pero no de NANOG, junto con KLF4 y cMYC, en células somáticas produce la reprogramación celular produciendo células inducidas pluripotentes (iPS) [86]. Estudios posteriores han revelado que KLF4 funciona conjuntamente con SOX2 y OCT4 potenciando sus efectos mientras que c-MYC controla otro tipo de procesos manteniendo la proliferación y el metabolismo [87]. Esto representa un gran avance no solo en cuanto al estudio de la pluripotencia, sino también para la medicina. Actualmente el uso de iPS es clave en el desarrollo de nuevos fármacos, en el estudio de diversas patologías desplazando a modelos animales y en medicina regenerativa [88].

6. Isoformas y *splicing* alternativo.

Una misma proteína puede tener distintas formas denominadas isoformas. Las isoformas se pueden generar a partir de genes independientes que forman parte de una misma familia y que se expresan independientemente en determinado momento o tejido, como por ejemplo la actina, o a partir del mismo gen por un mecanismo de *splicing* alternativo [89].

El *splicing* alternativo es un proceso común en células humanas donde se ha estimado que entre el 90% y el 95% de los genes multi-exón producen transcritos alternativos [90]. Las posibilidades son múltiples y se han definido diversos patrones de *splicing* alternativo (Figura 5). Uno de estos patrones corresponde a la utilización de un primer y último exón alternativo, de tal forma que la isoforma comienza o finaliza con una secuencia distinta. También algunas isoformas pueden saltar un exón sin incorporarlo a su secuencia. Otros exones son mutuamente excluyentes por lo que la utilización de uno anula la incorporación de otro. Por último, se puede retener un intrón como parte de un exón o utilizar sitios de 5' y 3' de *splicing* alternativos [91]. Esta multitud de opciones nutre a las distintas isoformas de

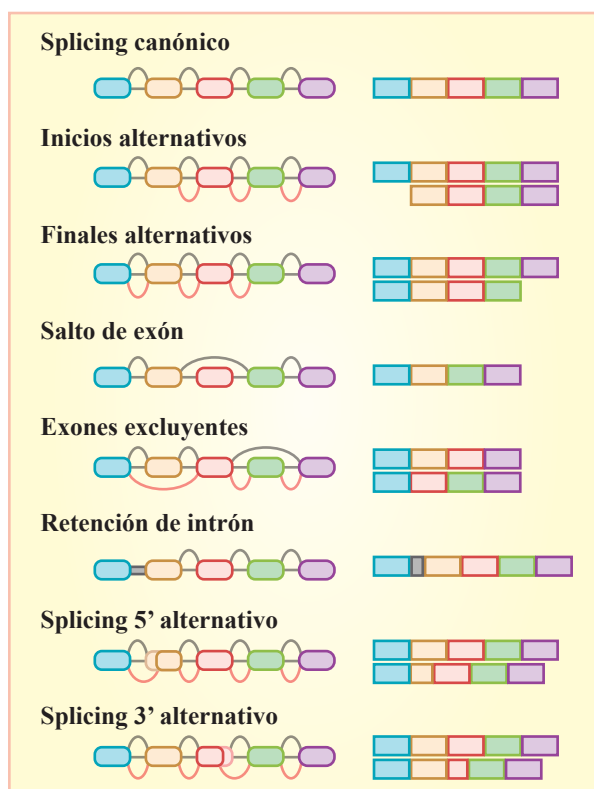


Figura 5. Patrones de *splicing* alternativo. Esquema de los distintos mecanismos de *splicing* alternativos. Esquema adaptado de Blencowe 2006.

características específicas que se pueden reflejar en aspectos como su localización celular o función.

Aunque un mismo gen puede dar lugar a varios transcritos que se traducirán en diferentes isoformas, análisis de secuenciación de ARN (RNA-Seq) han demostrado que la mayoría de los genes tienen un transcrito mayoritario que contribuye casi en un 85% al total del mRNA que codifica un loci y suele encontrarse expresado en varios tejidos indicando la expresión de una isoforma mayoritaria [92]. Sin embargo, se han detectado proteínas específicas de tejidos que pueden diferir por ejemplo en su localización celular o función [93]. Es por ello que la generación de isoformas, y por tanto el *splicing*, es un proceso importante en procesos como desarrollo, diferenciación e identidad tisular y cuya desregulación en muchas ocasiones está asociada con el desarrollo de ciertas patologías [94] [95]. Por citar un ejemplo, en cáncer se han observado alteraciones en la expresión de muchas isoformas asociadas al desarrollo de esta enfermedad [96] [97].

7. *Transcription Factor Like 5* (TCFL5).

Transcription Factor Like 5 (TCFL5) es un factor de transcripción de la familia bHLH descrito por primera vez en testículo, concretamente en espermatoцитos primarios en el estadio de paquitene [98]. En humanos, *TCFL5* se encuentra en el brazo q13.33 del cromosoma 20 en la hebra antisentido de la región cromosómica chr20: 62,841,115-62,861,763. Está compuesto por 6 exones y, únicamente la región bHLH, comprendida entre los exones 4 y 5, se conoce como dominio funcional de esta proteína. Hasta el momento, en la literatura se han demostrado fehacientemente la expresión de solo dos isoformas para este gen: TCFL5 y CHA (Figura 6). *TCFL5* está compuesto por 6 exones y presenta un marco de lectura abierto de 1500pb que codifica para una proteína de 500 aminoácidos. Por otro lado, *CHA* carece del primer exón por lo que comienza su expresión a partir del segundo presentando un marco de lectura abierto de 818pb que codifica para una proteína de 272 aminoácidos. Ambas isoformas expresan la región bHLH por lo que su única diferencia radica en la presencia o ausencia del primer exón.

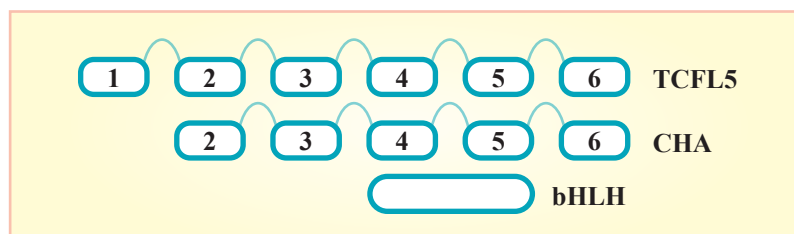


Figura 6. Representación de las isoformas TCFL5 y CHA. Se representan los exones que conforman las isoformas TCFL5 y CHA y la región bHLH. Esquema modificado de Rodríguez *et al*, 2003.

7.1. Funciones de las isoformas TCFL5 y CHA.

El papel de la isoforma TCFL5 ha sido descrito principalmente en espermatogénesis donde se sugiere un papel importante de dicho factor de transcripción. Durante este proceso, TCFL5 se une al promotor de la calmegina (CLGN), una chaperona requerida para procesos de espermatogénesis

y fertilidad, uniéndose a la región E-box CACGCG. La expresión de este factor de transcripción se encontró específica de espermatocitos primarios siendo su máxima expresión en estadio de diplotene de la profase meiótica donde colocaliza con la cromatina activa [99]. Posteriormente se describió su implicación en etapas más tardías en el proceso de espermiogénesis, concretamente en espermátidas, pero no en espermatozoides maduros. En espermátidas tempranas y tardías TCFL5 colocaliza con α -tubulina y con los centriolos formando la *manchette* que dará lugar al flagelo de los espermatozoides maduros. Además, TCFL5 interacciona con la proteína MEIG1 cuya delección genera la pérdida de fertilidad en ratón [100].

Por su parte, la isoforma CHA fue descrita por primera vez como autoantígeno de la enfermedad de Chagas. CHA presenta dos regiones de homología con proteínas del parásito *Trypanosoma cruzi*. La infección con *T. cruzi* produce una reacción cruzada entre la isoforma CHA y las proteínas del parásito lo que genera una respuesta autoinmune de linfocitos T y B provocando problemas cardíacos [101]. Además, únicamente los sueros de los pacientes chagásicos presentan título de anticuerpos contra CHA por lo que se ha descrito su uso como biomarcador para esta enfermedad [102]. A nivel molecular, CHA ha sido descrito como represor de la expresión de CD2 en linfocitos T. Esta represión se lleva a cabo a nivel transcripcional interaccionando con USF1 y uniéndose al promotor de CD2 en células T en reposo, pero no en células estimuladas [103].

Además, se ha atribuido también, un papel al gen TCFL5 en distintos procesos biológicos como activación y desarrollo del sistema inmune, cáncer o desarrollo embrionario sin llegar a discernir específicamente, en la mayoría de los casos, entre las isoformas TCFL5 y CHA.

En el sistema inmune se ha observado que TCFL5 es una diana transcripcional de NOTCH1 [104]. NOTCH1 es una proteína transmembrana que, tras la unión a sus ligandos, es procesado mediante dos cortes proteolíticos primero con una metaloproteasa y luego con un γ -secretasa. Este procesamiento libera en la célula la región intracelular de NOTCH1 (ICN) y se produce su translocación al núcleo, su unión con el complejo transcripcional CSL y activación de sus genes diana [105]. La inhibición del procesamiento de NOTCH1 reduce la expresión de TCFL5 en linfocitos T. Además, TCFL5 tiene un papel en el desarrollo de los linfocitos T en el timo generando un aumento de las células dobles positivas [104]. En células B, la activación de los receptores *Toll Like Receptor* 7 y 9 (TLR7 y TLR9) conllevan una inducción del mensajero de *TCFL5/CHA* [106]. Durante la diferenciación de monocitos, la expresión de *TCFL5/CHA* se ve reducida. Además, las proteínas *Retinoid X Receptor Beta* (RXRB) y *Core-Binding Factor Subunit Beta* (CBFB) se unen al promotor del gen *TCFL5* [107].

Por último, también se ha descrito a TCFL5 asociado negativamente a la infección del virus de la inmunodeficiencia humana (VIH) en nódulos linfáticos [108], asociado a la regulación por el factor de transcripción zinc finger PEG3 y por el bHLH ARNT2 en próstata [109] y en diferencias

de expresión en cerebro, donde se encuentra una mayor expresión en el ganglio estrellado asociado a sexo [110].

7.2. TCFL5 en cáncer.

Diversos estudios han mostrado superficialmente el papel de TCFL5 en cáncer de colon. Se ha descrito una inducción de *TCFL5/CHA* durante la formación de esferoides multicelulares tumorales (MCTS) [111], un modelo de agregación celular *in vitro* que simula la formación de un tumor [112] [113] [114]. El silenciamiento del gen conlleva una reducción del crecimiento de líneas celulares de cáncer de colon, del número de MCTS, así como un aumento de la viabilidad celular [111] [115]. Además, la expresión de *TCFL5/CHA* es mayor en carcinomas que en adenocarcinomas y está relacionada con la amplificación cromosómica y se ha postulado como uno de los genes importantes en la inestabilidad cromosómica relacionada con la progresión adenoma-carcinoma [116].

En leucemias, *TCFL5/CHA* se encuentra inducido en muestras de leucemias linfoblásticas agudas infantiles tanto T como B (T-ALL y B-ALL, respectivamente) resistentes al tratamiento con vincristina [117] [118]. En muestras infantiles B-ALL que presentan la translocación cromosómica t12:21, el promotor del gen *TCFL5* se encuentra hipometilado significativamente con respecto al resto de tipos de leucemias correlacionándose con un aumento en la expresión de este gen [119] [120]. Por último, *TCFL5/CHA* se encuentra diferencialmente inducido en seminomas espermatocíticos [121].

7.3. TCFL5 en pluripotencia y desarrollo embrionario.

El papel del TCFL5 en desarrollo embrionario no ha sido prácticamente estudiado. Solo a través de análisis multiómicos se ha podido definir a TCFL5 como uno de los genes cuya expresión se encuentra enriquecida en mESC [122] y que la deficiencia en la expresión de POU5F1 (OCT4) disminuye la expresión de este [123]. La deficiencia de la proteína MAGEB1 reprime la expresión de *Tcf5* en días tardíos de diferenciación [124]. Por otra parte, la sobreexpresión de TCFL5 en fibroblastos junto con los factores de transcripción tipo zinc finger ZFP238 y MYT1L favorece la diferenciación a células neuronales [125].

Objetivos

Objetivos

El factor de transcripción TCFL5 fue descrito por primera vez en testículo por Maruyama y colaboradores en 1998. Actualmente, más de 20 años después, su expresión y función sigue siendo una incógnita. Los trabajos publicados durante estos años acerca de TCFL5, ya sea en humano o en ratón, son escasos, aunque revelan diversos procesos biológicos en los que este gen podría estar implicado. Estos trabajos, en la mayoría de los casos, no describen a fondo una función para este factor de transcripción y tampoco discriminan entre las isoformas descritas. Sin embargo, se podrían definir 4 procesos biológicos en los cuales TCFL5 puede ejercer un papel relevante: espermatogénesis, sistema inmune, desarrollo embrionario y cáncer.

En el presente trabajo quisimos contribuir al conocimiento de este factor de transcripción. En consecuencia, el objetivo general describir las isoformas para TCFL5, así como estudiar la función de estas en procesos tumorales y desarrollo embrionario. Para ello, planteamos los siguientes objetivos específicos:

1. Describir a nivel experimental las distintas isoformas del gen *TCFL5*, así como su expresión tisular tanto en humanos como en ratón.
2. Estudiar el papel de las distintas isoformas de TCFL5 en procesos tumorales de cáncer colorrectal a nivel fenotípico y molecular.
3. Analizar el papel de TCFL5 en la regulación de genes de pluripotencia.
4. Estudiar la expresión de TCFL5 durante el periodo de preimplantación del desarrollo embrionario de ratón y su papel en procesos de diferenciación.

Materiales y métodos

Materiales y métodos

1. Materiales.

1.1. Líneas y cultivos celulares.

1.1.1. Líneas de cáncer de colon.

Las líneas celulares HT29-D6-luc, HT29, HCT116, SW620 y SW480 se crecieron en medio de cultivo MEM (*Minimun Essential Medium Eagle*) con 5% de suero fetal bovino (FBS) mientras que, la línea celular HEK293, se creció en medio DMEM (*Dulbecco's Modified Eagle Medium*) con 10% de FBS. Todos los medios de cultivo fueron suplementados con aminoácidos no esenciales, 0,01% de piruvato sódico, 2mM de glutamina, 100 U/ml de penicilina y 100 µg/ml de estreptomicina. El medio de cultivo fue renovado cada 48h y los pases se realizaron levantando las células con EDTA-Tripsina previo lavado con PBS estéril. Todo los compuestos y medios de cultivos fueron suministrados por el Servicios de Cultivos Celulares del Centro de Biología Molecular Severo Ochoa (CBMSO, Madrid, España), exceptuando el FBS que fue adquirido a las casas comerciales Gibco y Biowest. Los cultivos en monocapa se mantuvieron en un incubador bajo condiciones controladas de temperatura (37°C), atmósfera humificada (97%) y CO₂ (5%). Todo material plástico utilizado para el mantenimiento y desarrollo experimental de los cultivos celulares fue suministrado por Falcon (Beckton Dickson) y Corning (Corning Inc).

La línea HT29 (células epiteliales de adenocarcinoma de colon humano) se obtuvieron de la ATCC (LGC Standards, Barcelona, España). La línea HCT116 (células epiteliales de carcinoma de colon humano) se obtuvieron del Banco de células del Centro de Investigaciones Biológicas (CIB, Madrid, España). Las líneas SW620 (células de carcinoma de colon derivadas de metástasis) y SW480 (células de carcinoma de colon) fueron cedidas por el Dr. Alberto Muñoz del Instituto de Investigaciones Médicas (IIB, Madrid, España). La línea HEK293 (células epiteliales humanas de riñón) se encontraban disponibles en el CBMSO. Por último, la línea HT29-D6-luc (Xenogen Corporation) expresa constitutivamente el gen de la luciferasa cuya expresión se mantuvo estable tras la selección con 200 µg/ml del antibiótico G418.

1.1.2. Cultivo de MCTS.

Las líneas celulares se cultivaron en placas de baja adhesión celular a una densidad de 4×10^4 células por pocillo en medio de cultivo MEM suplementado con 0.4% FBS, 1 ng/ml de factor de crecimiento fibroblástico básico (bFGF) (Sigma), 10 ng/ml de factor de crecimiento epidérmico (EGF) (Sigma), 5 µg/ml de insulina (Sigma) y suplemento B27 (Invitrogene) a concentración final 1X. Los cultivos se mantuvieron durante 7 y 15 días dependiendo del experimento. Para reemplazar

el medio, los MCTS fueron centrifugados a 240 x g durante 5 minutos a temperatura ambiente. Posteriormente se volvieron a resuspender suavemente para no romperlos en medio fresco. Los cultivos de MCTS se mantuvieron en un incubador bajo condiciones controladas de temperatura (37°C), atmósfera humificada (97%) y CO₂ (5%).

1.1.3. Líneas embrionarias de ratón.

La línea embrionaria de ratón G4 (mESC) fue cultivada con medio DMEM 15% FBS en placas previamente tratadas con gelatina 0.2% (Sigma). El medio se suplementó con antibióticos, glutamina, aminoácidos no esenciales, β-mercapto-etanol, factor inhibidor de leucemia (LIF), 1 μM del inhibidor de MEK PD0325901 y 5 μM del inhibidor GSK3β CHIR99021. El medio fue renovado cada día y los pases se realizaron con EDTA-Tripsina tras 2 lavados con PBS estéril. Los compuestos, materiales y medios fueron suministrados por el laboratorio del Dr. Periklis Pantazis. Las células se mantuvieron en un incubador bajo condiciones controladas de temperatura (37°C), atmósfera humificada (97%) y CO₂ (5%).

Todos los experimentos con líneas embrionarias de ratón se llevaron a cabo en el laboratorio del Dr. Periklis Pantazis y bajo la supervisión de la Dra. Maaïke Welling en el departamento de Ciencias e Ingeniería de Biosistemas (D-BSSE, Basilea, Suiza) de la universidad *Eidgenössische Technische Hochschule* de Zurich (ETH Zurich, Suiza).

1.2. Cultivo de embriones in vitro.

Ratones hembras C57Bl/6 (Janvier Labs) fueron superovulados por tratamiento hormonal con gonadotropina y cruzados con ratones macho C57Bl/6 (Janvier Labs). Inmediatamente tras la cópula, las hembras fueron sacrificadas por inhalación de CO₂. Los embriones en estadio de 1 o 2 células fueron extraídos por lavado de los oviductos con medio M2 (Millipore) y cultivados en cámaras Lab-Tek (Nunc) a 37°C y 5% de CO₂ en medio KSOM con aminoácidos (Millipore) recubiertos por una gota de aceite mineral (Sigma). Los embriones se mantuvieron en cultivo hasta que alcanzaron el estadio de desarrollo deseado.

Todos los experimentos con embriones se llevaron a cabo en el laboratorio del Dr. Periklis Pantazis y bajo la supervisión de la Dra. Maaïke Welling y los responsables de los Servicios de Animalario en el departamento de Ciencias e Ingeniería de Biosistemas (D-BSSE, Basilea, Suiza) de la universidad *Eidgenössische Technische Hochschule* de Zurich (ETH Zurich, Suiza).

1.3. Plásmidos de expresión.

Plásmido	Característica	Uso
pCRII	Vector de fácil clonaje utilizado como intermedio en los procesos de clonaje. Presenta expresión de LacZ para selección de colonias por color (Invitrogen).	Clonaje intermedio de TCFL5 y CHA.
pCMV-Vsvg	Vector con la secuencia para la expresión de la envoltura viral (disponible en el laboratorio).	Generación de líneas estables.
pCgpV	Vector con la secuencia para la expresión de proteínas de empaquetamiento viral (disponible en el laboratorio).	Generación de líneas estables.
pLenti-CMV/TO-Hygro	Vector lentiviral con resistencia a higromicina B para su selección en células de mamífero (disponible en el laboratorio).	Clonaje de TCFL5 y CHA Generación de líneas estables.
pLenti-KO1-Puro	Vector lentiviral con resistencia a puomicina para su selección en células de mamífero (disponible en el laboratorio).	shARNs Generación de líneas estables.
pCDNA3-Flag	Vector de expresión bajo el control del promotor CMV. Presentan la etiqueta FLAG fusionada al gen de interés en el extremo C-terminal para mejorar el reconocimiento (OriGene Technologies).	Expresión de TCFL5 y CHA
pGL4	Vector con la expresión del gen luciferasa bajo el control de un promotor de interés a estudiar y de un promotor mínimo (disponible en el laboratorio).	Elemento de respuesta a NFkB2. Elementos de respuesta a SOX2, NANOG y KLF4. Promotor <i>SOX2_CORE</i> y <i>SOX2_R1*</i> .
pCas9	Vector con la expresión de la enzima Cas9 (ETH Zurich).	Generación de líneas deficientes en TCFL5
psgPal	Vector con una secuencia palindrómica y con capacidad de auto-escindir que recombina con las guías utilizadas en el sistema CRISPR-Cas9 (ETH Zurich).	Generación de líneas deficientes en TCFL5

Tabla 1. Plásmidos utilizados en los distintos experimentos. *Cedidos por el Dr. Ángel García Martín (StemTek Therapeutics)

1.4. Soluciones.

Nombre	Composición
Cristal Violeta	50% H ₂ O-metanol y 0,5% cristal violeta (p/v)
PBS / PBS-T	137 mM NaCl, 2,7 mM KCL, 8 mM Na ₃ PO ₄ y 1,5 mM K ₂ HPO ₄ / 0,1% Tween20 (v/v)
RIPA CHIP	10 mM Tris-HCl pH 8, 500 mM NaCl, 1% Triton X100 (v/v), 0,1% Na-deoxicolato, 0,1% SDS (v/v)
Sol. Concentradora (SDS-PAGE)	0,4 M Tris-HCl pH 6,8 y 0,4% SDS
Sol. Separadora (SDS-PAGE)	1,5 M Tris-HCl pH 8,8 y 0,4% SDS

Solución bloqueo	TBS-T o PBS-T más 5% BSA
Solución bloqueo IF	PBS más 10% suero de cabra
Solución electroforesis 10X	125 mM Tris-glicina, 25 mM Trizma base, 0,1% SDS pH8,3
Solución elución	50 mM Tris-HCl pH 8, 10 mM EDTA, 1% SDS (v/v)
Solución IP	1% Tritón X100 (v/v), 150 mM NaCl, 50 mM Tris-HCl pH 7,5, 2 mM EDTA, 1 mM EGTA
Solución fijación	4% Paraformaldehído (PFA)
Solución lisis CHIP (LB1)	50 mM Hepes-KOH pH 7,5, 150 mM NaCl, 1 mM EDTA, 10% glicerol (v/v), 0,5% NP40 (v/v), 0,25% Triton X100 (v/v)
Solución lisis CHIP (LB2)	10 mM Tris-HCl pH 8, 200 mM NaCl, 1mM EDTA, 0,5 mM EGTA
Solución lisis CHIP (LB3)	10 mM Tris-HCl pH 8, 100 mM NaCl, 1mM EDTA, 0,5 mM EGTA, 0,1% Na-deoxicolato, 0,1% SDS (v/v)
Solución lisis de proteínas	1% Tritón X100 (v/v), 400 mM NaCl, 20 mM Tris-HCl pH 8, 1 mM EDTA, 1mM EGTA
Solución lisis Proteinasa K	1M Tris-HCl pH 8, 5M NaCl, 0.5M EDTA, 10% SDS (v/v)
Solución TE	10 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM EDTA
Solución tinción	PBS más 5% suero de cabra
Solución transferencia 10X	25 mM Tris-HCl pH 8,3, 192 mM glicina. Al uso añadir 20% metanol
Solución X-Link Reverso	50 mM Tris-HCl pH 8, 0,3 M NaCl, 10 mM EDTA, 1% SDS (v/v)
TAE	40 mM Trizma base, 20 mM Ac. Acético y 1 mM EDTA pH 8
TBS / TBS-T	20 mM Tris-HCl pH 7,6 y 137 mM NaCl / 0,05% Tween20 (v/v)

Tabla 2. Soluciones y tampones utilizados en los distintos experimentos.

1.5. Bases de datos.

Nombre	Uso	Sitio web
cBioPortal	Expresión	www.cbioportal.org/
CHIP Atlas	CHIP-Seq	www.chip-atlas.org/
ECGene	Secuencias	www.genome.ewha.ac.kr/ECgene/
Ensemble	Secuencias	www.ensembl.org/index.html
NCBI	Secuencias Alineamientos Referencias Expresión RNA-Seq	www.ncbi.nlm.nih.gov/
OncoMine	Expresión	www.oncomine.org/resource/login.html
RoadMap	Promotores	www.roadmapepigenomics.org/
UCSC Genome Browser	CHIP Seq	www.genome.ucsc.edu/
UniProt	Secuencias	www.uniprot.org/
Vega	Secuencias	www.vega.archive.ensembl.org/index.html

Tabla 3. Bases de datos consultadas para el desarrollo experimental.

1.6. Anticuerpos.

Anticuerpo	Casa Comercial	Referencia	Uso* (Dilución)
Anti-FLAG	Sigma Aldrich	F1804	WB (1:1000) IP (0.2 µg)
Anti-Mouse IgG / HPR	Cell Signaling	7076S	WB (1:2000)
Anti-Mouse IgG / Alexa 594	ThermoFisher	A-21203	C (1:500)
Anti-Rabbit IgG / HPR	Cell Signaling	7074S	WB (1:5000)
Anti-Rabbit IgG / Alexa 488	ThermoFisher	A21202	IF (1:500)
CD133-APC	ThermoFisher	17-1338-42	C (1:100)
CDK4	Cell Signaling	12790S	WB (1:1000)
EPHB2	ThermoFisher	MA5-11162	C (1:50)
HSP90	Santa Cruz Biotechnology	Sc13119	WB (1:1000)
LGR5	Abcam	Ab75850	C (1:100)
NANOG	Cell Signaling	4903P	WB (1:1000)
NFKB2	Cell Signaling	4882S	WB (1:1000) IP (1:50)
p21	Cell Signaling	2947S	WB (1:1000)
p27	Cell Signaling	2552S	WB (1:1000)
SOX2	Cell Signaling	3579S	WB (1:1000)
TCFL5	Sigma Aldrich	HPA055223	WB (1:2000) IF (1:500)
TCFL5_E1**	GenScript	-	WB (1:1000) IF (1:1000)

Tabla 4. Anticuerpos primarios y secundarios utilizados durante el desarrollo experimental. *C: Citometría; IF: Inmunofluorescencia; IP: Inmunoprecipitación; WB: Western-blot. **Anticuerpo policlonal de conejo generado contra el péptido de TCFL5: AEPPAPRGPEPPE.

1.7. Oligonucleótidos.

1.7.1. Oligonucleótidos para uniones de exones.

Uniones de exones humano

Unión	Sentido	Antisentido
E1/E2	5'ACCGCTTCAACAGCATCC3'	5'GATGGATGTCGAATGAGAGTTACC3'
E2/E3	5'CCCACTGTTTACTACAAATGCT3'	5'CTCTAGGCAATCCAATATCCTG3'
E2b/E3	5'GACTTCCACAGCAGCGATAG3'	5'CTCTAGGCAATCCAATATCCTG3'
E3/E4	5'GAGCAAGTTTGGATTAAAGTGGG3'	5'ACTGACGCATTCTACTCCGA3'
E4/E5	5'GAGAGGCATAACCGAATGGA3'	5'GCAGAACGGCACTAAGAG3'
E5/E6	5'CACAGCATTCCTGAAATACATCC3'	5'GCAGAAGTATATTACCGAGTCC3'
E6/E7	5'GTATTCTGATCGGATCGTTTATGG3'	5'CATGACGTATCTGGAGTTTGTG3'

E7/E8	5'GAATCAGTTGAACCCAGGAG3'	5'GTCTGGTCAGCTTTAGCCT3'
E5/E7	5'CACAGCATTCTGAAATACATCC3'	5'GGATTCTTTCCATCAACTGCT3'
E5/E8	5'CACAGCATTCTGAAATACATCC3'	5'GTCTGGTCAGCTTTAGCCT3'

Uniones de exones ratón

Unión	Sentido	Antisentido
e1/e2	5'GACCGCTTCAACAGTATGCC3'	5'GTTACTAGAACTCTGCGTCTGTG3'
e2/e3	5'ACTCATGCTCTATACTGGAAGCTG3'	5'CTTCTCCCACCTTAATCCAAACCT3'
e1/e3	5'GACCGCTTCAACAGTATGCC3'	5'CTTCTCCCACCTTAATCCAAACCT3'
e3/e4	5'GCAGGTTTGGATTAAAGTGGGA3'	5'AAGAGCCTTCACCTACATTCTG3'
e3/e4b	5'GCAGGTTTGGATTAAAGTGGGA3'	5'ACAGCCAGGGCTACACATCT3'
e4/e5	5'CGGAGCAGGTTTGGATTAAA3'	5'GGGTTGTTGCTTTATCTGTCTC3'
e4/e6	5'CGGAGCAGGTTTGGATTAAA3'	5'TCAGTCCAGTCACTTGATCTCC3'
e5/e6	5'AGTACATTGAGGAAAGACATGGG3'	5'TCAGTCCAGTCACTTGATCTCC3'

Tabla 5. Oligonucleótidos utilizados para la amplificación de las uniones de exones de las isoformas de TCFL5 en humanos y en ratón**1.7.2. Oligonucleótidos para qPCR.****Oligonucleótidos qPCR - Humano**

Gen	Sentido	Antisentido
<i>C-MYC</i>	5'GCTGCTTAGACGCTGGATTT3'	5'TAACGTTGAGGGGCATCG3'
<i>GAPDH</i>	5'GCACAAGAGGAAGAGAGAGACC3'	5'AGGGGAGATTGAGTGTGGTG3'
<i>KLF4</i>	5'GTGGAGAAAGATGGGAGC3'	5'TGACTTTGGGGTTCAGGT3'
<i>NANOG</i>	5'TCTCCAACATCCTGAACCTCA3'	5'TTGCTATTCTTCGGCCAGTT3'
<i>POU5F1</i>	5'TCTTTCCACCAGGCCCCCGGCTC3'	5'TGCGGGCGGACATGGGGAGATCC3'
<i>SOX2</i>	5'GGTTCAGAAGCGAAAGCAGT3'	5'CCGCATCTCTTCCTCACG3'
<i>TCFL5/CHA</i>	5'CACAGCATTCTGAAATACATCC3'	5'GTCTGGTCAGCTTTAGCCT3'
<i>TCFL5_E7</i>	5'CACAGCATTCTGAAATACATCC3'	5'GGATTCTTTCCATCAACTGCT3'
<i>TCFL5_E6</i>	5'CACAGCATTCTGAAATACATCC3'	5'GCAGAAGTATATTCACCGAGTCC3'
<i>TP53</i>	5'GTGGAATGCAAGCCAATTCAG3'	5'GAGGAGTGTGGGTGCCTCT3'

Oligonucleótidos qPCR – Ratón

Gen	Sentido	Antisentido
<i>DAZL</i>	5'TCGAAGGGCTATGGATTTGT3'	5'ACGTGGCTGCACATGATAAG3'
<i>e3/e4</i>	5'CGGAGCAGGTTTGGATTAAA3'	5'CCGATTCTCTTATTGATTGCT3'
<i>e4b</i>	5'TTCCCAGACACAAGTGAAAGG3'	5'ACTAACCTGAAGGCCCAAGG3'
<i>e5/e6</i>	5'CAGGAAAGACATGGGGACTC3'	5'GGGATTGCGGTCTAGTCAGC3'
<i>GAPDH</i>	5'CCTGCTTCACCACCTTCTTG3'	5'CTGAGAACGGGAAGCTTGTC3'
<i>NANOG</i>	5'TTGGAAGCCACTAGGGAAAG3'	5'AGCCCAGATGTTGCGTAAGT3'
<i>POU5F1</i>	5'CAAGGCAAGGGAGGTAGACA3'	5'CACCCCTGTTGTGCTTTTAAT3'
<i>SOHLH2</i>	5'GGGCAGGGCAGAGTAAATCTT3'	5'CAAACGAGTTAGCAGCCAAAAG3'
<i>SOX2</i>	5'TCGAGATAAACATGGCAATCA3'	5'AACCCAGCAAGAACCCTTTC3'
<i>STELLA</i>	5'TTTGTTGTGCGGTGCTGAAAG3'	5'TCCCGTTCAAACCTCATTTC3'

Tabla 5. Oligonucleótidos utilizados para cuantificación de la expresión génica en humanos y en ratón.

1.7.3. Oligonucleótidos para inmunoprecipitación de cromatina (CHIP).

Oligonucleótidos CHIP

Promotor	Sentido	Antisentido
BCL2_01	5'TTCGCAGAAGTCCTGTGATG ^{3'}	5'GCATTTGCTGTTCCGGAGTTT ^{3'}
BCL2_02	5'CTTCACATGGCATGATCTGG ^{3'}	5'CTGGGTACCATTACAGGGAGA ^{3'}
KLF4 neg	5'TCTCCCTCCAGGAAAAGGAG ^{3'}	5'CCTGAAGCCTGCATTCTAGC ^{3'}
KLF4_01	5'ATTATCCGCGTGAATCATCC ^{3'}	5'CCTCCTCTCCACACCCCTA ^{3'}
KLF4_02	5'CTGCTCAGCACTTCCTCAAG ^{3'}	5'GTCTCTTCGTGCACCCACTT ^{3'}
SOX2 CORE_01	5'AGCTGAGTTGGACAGGGAGA ^{3'}	5'CAGGTGCCAACACTCTCTCA ^{3'}
SOX2 CORE_02	5'GCGTCCCATCCTCATTTAAG ^{3'}	5'AGCAACAGGTACACCCACAC ^{3'}
SOX2 R1_01	5'CACACTCCTCTTCCCCTTGA ^{3'}	5'GAGAATGCTTCCAGCCAGAG ^{3'}
SOX2 R1_02	5'GGTGGTCGTCAAACCTCTGCT ^{3'}	5'TTGAGTGTGTTCCCCCTCCTC ^{3'}
TCFL5_01	5'GGCAGGAATGAAGACGTTGT ^{3'}	5'CTCGATCTCCTGACCTCCTG ^{3'}
TCFL5_03	5'GTGATACCTACGCGGTGACA ^{3'}	5'CTTCATTCTGCGCGCTGTA ^{3'}
TCFL5_04	5'GAAAGCCCAGAAGTGTCAGC ^{3'}	5'GAGGAGGACTTCCACAGCAG ^{3'}
TCFL5_05	5'GGGAGGGGGTGTGTTAGTTTC ^{3'}	5'AAAGGCTGTGTGCGCTGT ^{3'}
TCFL5_06	5'GAATCAGGAATAGGCGGTGA ^{3'}	5'CTAAGCTTAAGCGGCCTGTG ^{3'}
TCFL5_07	5'TCAAGAGGCTGAGGCAGAAT ^{3'}	5'CCCAGGCTGAAGTGTAGTGG ^{3'}
TCFL5_08	5'CCAGATAACTGCCAGGGTGT ^{3'}	5'CCTCCCTCCCTCTACAGTCC ^{3'}

Tabla 6. Oligonucleótidos utilizados para el análisis de los ensayos de inmunoprecipitación de cromatina.

1.7.4. Oligonucleótidos para clonaje.

Oligonucleótidos Clonaje

Gen	Sentido	Antisentido
<i>CHA</i>	5'AAGGATCCATGAATGTTCTCTTCAGCAACAAAACAAATG ^{3'}	5'GTCTAGATCACTTGATCTCCATCGAG ^{3'}
<i>TCFL5</i>	5'CAAGGATCCATGTCGGGCCCCGGAC ^{3'}	5'GTCTAGATCACTTGATCTCCATCGAG ^{3'}

Tabla 7. Oligonucleótidos utilizados para la amplificación de las secuencias de ADNc de TCFL5 y CHA para su posterior clonaje. En rojo, la secuencia de la enzima de restricción BamHI. En azul, la secuencia de la enzima de restricción XbaI.

1.7.5. Secuencias de shARNs.

Secuencias shARNs

Código	Nombre	Secuencia
TRCN0000014612	shTCFL5_12	5'GAATCCACTAAACAGACGTTACTCGAGTAACGTCTGTTAGTGGATTC ^{3'}
TRCN0000418429	shTCFL5_29	5'TAGCAATTGGAACAAGTTAACTCGAGTTAACTTGTCCAATTGCTA ^{3'}
TRCN0000435190	shTCFL5_90	5'CGACATCCATCTGAACATAATGCTCGAGCATTAGTTCAGATGGATGTCG ^{3'}

Tabla 8. Oligonucleótidos utilizados para el silenciamiento de TCFL5 y CHA por shARNs. En rojo la secuencia sentido y antisentido de reconocimiento de la secuencia de TCFL5.

1.7.6. Secuencias CRISPR-Cas9.

Guías CRISPR Humano

Nombre	Secuencia + ProtoSpacer
E3 A	5'TGGAAAGGACGAAACACCACCTGGTTAATTTTCGTACGGTTTAAGAGCTATGCTGGAAAC ^{3'}
E3 B	5'TGGAAAGGACGAAACACCACCCCGTACGAAAATTAACCGTTTAAGAGCTATGCTGGAAAC ^{3'}
E3 C	5'TGGAAAGGACGAAACACCACCTGGTTAATTTTCGTACGGTTTAAGAGCTATGCTGGAAAC ^{3'}
E4 A	5'TGGAAAGGACGAAACACCACCTGGTTAATTTTCGTACGGTTTAAGAGCTATGCTGGAAAC ^{3'}
E4 B	5'TGGAAAGGACGAAACACC GTGTGAGAACGTTGCGTTTGGTTTAAGAGCTATGCTGGAAAC ^{3'}
E4 C	5'TGGAAAGGACGAAACACCTATGTTATACTCGATGAGTGGTTTAAGAGCTATGCTGGAAAC ^{3'}

Guías CRISPR Ratón

Nombre	Secuencia + ProtoSpacer
e3 A	5'GGAAAGGACGAAACACCTAGTAGAAACAAAGCCTTGCGTTTAAGAGCTATGCTGGAAAC ^{3'}
e3 B	5'GGAAAGGACGAAACACCCGAGTCCACCAACAGACAGGGTTTAAGAGCTATGCTGGAAAC ^{3'}
e3 C	5'GGAAAGGACGAAACACCCATTGAGTCCACCAACAGACGGTTTAAGAGCTATGCTGGAAAC ^{3'}
e4 A	5'GGAAAGGACGAAACACCTCGGGAGAGGCATAACCGAAGTTTAAGAGCTATGCTGGAAAC ^{3'}
e4b A	5'GGAAAGGACGAAACACCCACCAATGAGGCCATCATACTTTAAGAGCTATGCTGGAAAC ^{3'}
e4b C	5'GGAAAGGACGAAACACCCGTGAGCCATGTTCTCTGCACGGTTTAAGAGCTATGCTGGAAAC ^{3'}
e5 A	5'GGAAAGGACGAAACACCCAGCATTCTGAAGTACATTCGGTTTAAGAGCTATGCTGGAAAC ^{3'}
e5 B	5'GGAAAGGACGAAACACCCGTACATTACAGAAAGACATGGTTTAAGAGCTATGCTGGAAAC ^{3'}
e6 A	5'GGAAAGGACGAAACACCCGGCAGGTCACCAGGGATTCGGTTTAAGAGCTATGCTGGAAAC ^{3'}

Tabla 9. Secuencias de las guías utilizadas para la delección de TCFL5 en humanos y en ratón por la técnica *Self-cloning* CRISPR-Cas9. En roja la secuencia de la guía. En negro la secuencia complementaria para la amplificación por PCR.

1.7.7. Oligonucleótidos para amplificación de las guías para CRISPR-Cas9.

Amplificación Guías CRISPR

Nombre	Secuencia
sgRNA_HDRstep1_fw	5'TGTTTTAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTCGATTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACC ^{3'}
sgRNA_HDRstep1_rv	5'GTTGATAACGACTAGCCTTATTAACTTGCTATGCTGTTCCAGCATAGCTCTTAAC ^{3'}
sgRNA_HDRstep2_fw	5'GTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTAAATTATGTTTTAAATGGACTATCATATGCTTACC ^{3'}
sgRNA_HDRstep2_rv	5'ATTTTAACTTGCTATTTCTAGCTCTAAAACAAAAAAGCACCGACTCGGTGCACATTTTCAAGTTGATAACGGACTAGCCTTATTAAAC ^{3'}
sgRNA_HDRstep3_fw	5'CGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAA ^{3'}
sgRNA_HDRstep3_rv	5'TCAATGTATCTTATCATGTCTGCTCGATTTAACTTGCTATTTCTAGCTCTAAAACAAAA ^{3'}
sgRNA_HDRstep4_fw	5'GGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATA ^{3'}
sgRNA_HDRstep4_rv	5'TCAATGTATCTTATCATGTCTGCTCGA ^{3'}

Tabla 10. Oligonucleótidos utilizados para la amplificación de las guías para la técnica *Self-cloning* CRISPR-Cas9.

1.7.8. Oligonucleótidos secuenciación.

Oligonucleótidos secuenciación humano

Nombre	Secuencia	Región a secuenciar
Exón 2b	5'GACTTCCACAGCAGCGATAG ^{3'}	Exón 3
Exón 4	5'ACTGACGCATTCTACTCCGA ^{3'}	

Oligonucleótidos secuenciación ratón

Nombre	Secuencia	Región a secuenciar
Intron2_F	5'TGTTGAGTGCATTTTGCTCA ^{3'}	Exón 3
Intron3_R	5'ATGAGGAGTGCAGGGTGAAC ^{3'}	
Intron3_F_01	5'CGAAGAGCCCCTTGAGAGAT ^{3'}	Exón 4
Intron4_R_01	5'CAAAGTCAGCAAACGCAAA ^{3'}	
Intron3_F_02	5'TTTGCGTTTGCTGACTTTTG ^{3'}	Exón 4b
Intron4_R_02	5'TGGCCTTTCACCTGTGTCTG ^{3'}	
Intron4_F	5'CTCAGTGCTCAGTGGCAAAG ^{3'}	Exón 5
Intron5_R	5'TGCCAATCACATCGAGAATC ^{3'}	
Intron5_F	5'GAGATAGCCACCAACCTGGA ^{3'}	Exón 6
Intron6_R	5'CATCCCTCGAGAGGACTCAG ^{3'}	

Tabla 11. Oligonucleótidos utilizados para la secuenciación de las líneas deficientes en la expresión de TCFL5 y CHA.

2. Métodos.

2.1. Alineamiento de secuencias.

Los métodos de carácter bioinformático se realizaron en el Servicio de Genómica y Secuenciación Masiva del Centro de Biología Molecular Severo Ochoa (CBMSO, Madrid, España) por el Dr. Alberto Rastrojo. Brevemente, se alinearon todas las secuencias anotadas en bases de datos (Tabla 3) para TCFL5 en humanos y en ratón para la generación de secuencias consenso. Los distintos datos de RNA-Seq para la descripción de las isoformas y uniones de exones se obtuvieron del repositorio *Sequence Read Archive* (SRA, NCBI).

2.2. Transfección transitoria.

Se sembraron las líneas celulares 24 horas antes de la transfección en placas de 24, 12, 6 pocillos o placas p100 dependiendo del experimento a un 70% de confluencia en medio completo sin antibiótico. Las transfecciones de ADN se realizaron mediante partículas lipídicas utilizando Metafectene Pro (Biontex) siguiendo las instrucciones del fabricante. Los plásmidos se prepararon de manera independiente a las concentraciones indicadas por el fabricante en medio libre de suero Opti-MEM (Gibco). El medio de cultivo se reemplazó a las 24 h del inicio de la transfección. En las siguientes 24-72 h se utilizaron las células transfectadas para los diferentes análisis requeridos para cada experimento.

2.3. Clonaje y crecimiento bacteriano.

La secuencia de las isoformas TCFL5 y CHA se obtuvieron mediante amplificación por PCR con oligonucleótidos específicos para cada isoforma (Tabla 8) del vector pCMV6-XL5-TCFL5 (OriGene). El producto de PCR fue purificado del gel de agarosa con el kit QIAquick Gel Extraction Kit (Qiagen), clonado en el vector pCRII y transformado en bacterias *Escherichia coli* (IVNαF') utilizando el kit Strategene "TA cloning kit" (Invitrogen) siguiendo las instrucciones de los fabricantes. Las bacterias transformadas se sembraron en placas Petri en medio LB-agar, 50 µg/ml de Kanamicina y 40 mg/ml de X-Gal para la selección por color (colonias blancas/azules). Las colonias blancas aisladas se crecieron en 5 mL de medio LB y 50 µg/ml de Kanamicina durante 16 h a 37°C. La extracción y purificación de los plásmidos pCRII clonados se realizó utilizando el kit Spin Miniprep kit (Qiagen) según las instrucciones del fabricante. Los plásmidos aislados se resolvieron en gel de agarosa 1% previa digestión con las enzimas de restricción BamHI y XbaI (New England Biolabs) y se mandaron a secuenciar al servicio de genómica del parque científico de Madrid para corroborar la secuencia. Previamente se realizó una digestión del vector pLenti-CMV/TO-Hygro con las mismas enzimas de restricción. Se extrajo y se purificó el ADNc de las isoformas TCFL5 y CHA tras la digestión del plásmido pCRII, así como el vector pLenti-CMV/TO-Hygro digerido y se ligaron con la enzima T4 ligasa (New England Biolabs) siguiendo el protocolo del fabricante. El producto de la ligación se transformó en bacterias *E. coli* (DH5α) competentes y se crecieron en placas Petri con LB-agar y 100 µg/mL de ampicilina. Se crecieron las colonias resultantes, se purificaron y se resolvieron en gel de agarosa 1% tras la digestión con las enzimas de restricción BamHI y XbaI para comprobar la correcta ligación. Una vez corroborado el correcto clonaje del plásmido, se crecieron las bacterias transformadas en 500 mL de LB con 100 µg/mL de ampicilina durante 16 h a 37°C y se obtuvo el stock final tras la purificación del plásmido utilizando el kit Plasmid Maxi Kit (Qiagen) siguiendo las instrucciones del fabricante.

2.4. Generación de líneas estables por lentivirus.

La generación de las líneas celulares HT29-D6-luc, HT29, HCT116 y SW620 con la sobreexpresión estable de las isoformas TCFL5 y CHA y el silenciamiento por ARN de interferencia (shARN) se realizaron mediante partículas lentivirales generadas por el sistema ViralSafe Lentiviral Bisistronic Expression System (Cell Biolabs) siguiendo las instrucciones del fabricante.

Las partículas virales se generaron en células HEK293 por transfección transitoria de los vectores pCMV-Vsvg y pCspV, junto con los vectores pLenti-CMV/TO-Hygro, pLenti-CMV/TO-Hygro-CHA o pLenti-CMV/TO-Hygro-TCFL5 (previamente clonados) para las líneas sobreexpresadas, o los vectores pLenti-KO1-puro y pLenti-KO1-shTCFL5_12, pLenti-KO1-shTCFL5_29 o pLenti-KO1-shTCFL5_90 (OriGene) para las líneas silenciadas. El medio de cultivo de las transfecciones que contienen las partículas virales se recogió a las 24 y 48h tras la transfección.

La infección con los lentivirus se realizó añadiendo el medio viral filtrado con filtros de nitrocelulosa de 0.45 μm y medio completo en proporción 1:1. Esta mezcla se añadió a las células HT29-D6-luc, HT29, HCT116 y SW620 en un 50% de confluencia, junto con 8 $\mu\text{g/ml}$ de polibreno para aumentar la eficiencia de infección. Transcurridos 2 días del último ciclo de infección, las líneas celulares HT29-D6-luc, HCT116 y SW620 generadas para la sobreexpresión del vector vacío (EV), CHA o TCFL5, se seleccionaron añadiendo al medio 200 $\mu\text{g/ml}$ de higromicina B (Invitrogene) para eliminar aquellas células que no han sido infectadas. Las líneas celulares HT29 y HCT116 silenciadas shTCFL12, shTCFL5_29 y shTCFL5_90 se seleccionaron con 4 $\mu\text{g/ml}$ de puomicina (Invitrogene). La selección se renovó diariamente durante aproximadamente una semana. Finalmente, la expresión de las líneas celulares generadas fue analizada por RT-qPCR y WB.

2.5. Generación de líneas estables por edición génica (*Self-cloning* CRISPR-Cas9).

Para la generación de líneas *knock-out* se utilizó la técnica *Self-cloning* CRISPR-Cas9 modificando en cada caso el protocolo descrito por Arbab y colaboradores [126]. Esta técnica aprovecha la capacidad de reparación del daño en el ADN de las células embrionarias y cancerígenas por recombinación homóloga y permite la generación de líneas *knock-out* de manera rápida y eficiente. La técnica utiliza un plásmido que contiene la enzima Cas9 (pCas9), un plásmido palindrómico (sgPal) con capacidad de auto-escindirse, que genera las sgRNAs en el interior de la célula, y las sgRNAs específicas generadas por PCR y complementarias al sgPal. De este modo, al transfectar todos los elementos en la célula, primero se genera un sgRNA por el plásmido sgPal que, junto con la Cas9, produce un daño en el propio sgPal. Después, ese daño en el plásmido sgPal es reparado por recombinación homóloga con las sgRNAs generadas por PCR. Por último, el plásmido sgPal recombinado genera las nuevas sgRNA específicas y actúa junto con la Cas9 para producir la edición génica en el gen deseado.

2.5.1. Líneas embrionarias de ratón.

Para la generación de líneas mESC G4 deficientes en la expresión de TCFL5, se siguió el protocolo descrito por Arbab y colaboradores [126]. Brevemente, las guías (Tabla 10) fueron amplificadas en 3 PCRs para generar la secuencia de ADN que contiene la guía de reconocimiento y la zona de recombinación del plásmido psgPal. Las reacciones de PCR se llevaron a cabo utilizando la enzima Q5 *High Fidelity* (New England Biolabs) en el termociclador T100 Thermal cycler (BioRad) siguiendo las instrucciones del fabricante. La primera y segunda PCR se realizaron con 10 ciclos de amplificación mientras que la tercera se realizó con 35 ciclos de amplificación. Posteriormente se resolvió el producto final de amplificación en gel de agarosa 1% y se purificó utilizando el kit QIAquick Gel Extraction Kit (Qiagen). Una cantidad de 10^6 mESC G4 fueron electroporadas con 10 μl de guía purificada, 4 μg de plásmido psgPal y 4 μg de pCas9 bajo las condiciones 230 V, 05 mF y máxima resistencia. Tras 24 horas, las células fueron seleccionadas con 10 $\mu\text{g/ml}$ de blasticidina y

100 µg/ml de higromicina durante 48 horas. Las células supervivientes fueron cultivadas en placas de 96 pocillos a densidad de 1 célula por pocillo tras su separación celular por sorter en FACSVantage SE (BD Bioscience). Los clones que se generaron fueron secuenciados (Tabla 12) y validados por qPCR..

2.5.2. Línea celular HCT116.

Para la generación de líneas HCT116 deficientes en la expresión de TCFL5, se modificó el protocolo descrito. Se utilizaron dos guías flanqueantes del exón E3 de TCFL5 para su escisión total (Tabla 10). Estas guías fueron amplificadas de igual forma que la descrita anteriormente. Una cantidad de 3×10^5 células HCT116 se transfectaron con 10 µl de la mezcla de dos guías flanqueantes purificadas, 0,5µg de plásmido psqPal y 0,5µg de pCas9 utilizando Metafectene Pro (Biontex) siguiendo las instrucciones del fabricante. Tras 48 horas, se seleccionaron las células con 10 µg/ml de blasticidina y 200 µg/ml de higromicina durante 48 horas. De igual manera las células supervivientes fueron cultivadas en placas de 96 pocillos a densidad de 1 célula por pocillo tras su separación celular por sorter en FACSVantage SE (BD Bioscience). Los clones que se generaron fueron secuenciados (Tabla 12) y validados por PCR y qPCR.

2.6. Diferenciación de mECS a línea germinal.

Las mESCs G4 generadas se sembraron a confluencia en placas de 24 pocillos tratadas con 0.2% de gelatina de tal forma que cada pocillo representase un punto del experimento. Las mESCs se mantuvieron durante 10 días en medio IMDM (Invitrogen) completo suplementado con 15% FBS inactivado, 25 µg/ml de ácido ascórbico (Sigma), 1-tioglicerol (Sigma) y 0.2 mg/ml de holo-transferrina (Sigma). El medio se reemplazó a día 2 y después a diario hasta la finalización. Cada dos días se obtuvo un punto para el posterior análisis por qPCR.

2.7. Obtención y análisis de ADN.

2.7.1. Extracción.

El ADN se extrajo a partir de células cultivadas in vitro. Las células se incubaron en 500 µL de solución de lisis para Proteinasa K durante 12 h a 55°C. Posteriormente, se incubó con 300 µL de solución NaCl saturada 10 minutos a 4°C y se centrifugó a 17900 x g durante 20 minutos a 4°C. El sobrenadante se incubó con 800 µL de isopropanol durante 15 minutos a 4°C y se centrifugó a 17900 x g durante 20 minutos a 4°C. El pellet se lavó con 1 mL de etanol al 70% y se centrifugó a 17900 x g durante 5 minutos a 4°C. Por último, el pellet se resuspendió en 50 µL de solución TE y se incubó la solución de DNA a 55°C durante 15 minutos. El ADN obtenido se cuantificó y se determinó su grado de pureza con un espectrofotómetro Nanodrop 1000 (Thermo Scientific).

2.7.2. Secuenciación.

El ADN obtenido de células fue enviado a secuenciar al Servicio de Genómica del Parque Científico de Madrid con oligonucleótidos específicos (Tabla 12) siguiendo sus recomendaciones. Los resultados de la secuenciación fueron visualizados en el programa Chromas DNA Sequencing Software (Technelysium).

2.7.3. Inmunoprecipitación de cromatina (CHIP).

Para la inmunoprecipitación de cromatina (CHIP), 1×10^7 células fueron utilizadas por cada condición. Las células se fijaron con 1% de formaldehído durante 10 minutos a temperatura ambiente y posteriormente se añadió glicina 2.5 M en proporción 1/20. Se lavaron las células 2 veces con PBS frío y el pellet de células fue congelado inmediatamente a -80°C . A continuación, se procedió al lisado de las células. El pellet de células fue incubado en 1 mL de solución de lisis LB1 durante 10 minutos a 4°C y se centrifugó a $1700 \times g$ durante 5 minutos a 4°C . El pellet se resuspendió en 1 mL de solución de lisis LB2, se incubó 5 minutos a 4°C y se centrifugó a $660 \times g$ durante 5 minutos a 4°C . Por último, el pellet se resuspendió en 1 mL de solución de lisis LB3 y se procedió a la sonicación de la muestra en Bioruptor Next Gen (Diagenode). Se utilizaron 10 ciclos de 30 segundos de sonicación y 30 segundos de parada. Para comprobar la longitud del ADN entre 300 y 400pb, se recogieron 20 μL de la muestra y se añadieron 3 volúmenes de solución X-Link, se incubó a 65°C durante 12 h y se resolvió en gel de agarosa al 1%. Tras la comprobación del ADN, se añadió Triton X100 y NP40 para una concentración final de 1% y 0.5%, respectivamente. Se centrifugó el lisado a $20800 \times g$ durante 10 minutos a 4°C . El equivalente a 5×10^5 células se guardó como muestra control. El resto se utilizó para la inmunoprecipitación de la cromatina.

El lisado celular se incubó 12 h con bolas magnéticas fusionadas al anticuerpo α -FLAG (Sigma) a 4°C en rotación. La purificación de las bolas magnéticas se llevó a cabo utilizando el imán QuadroMACS Separator (Miltenyl Biotec). Las bolas magnéticas se lavaron 5 veces en 1 mL de RIPA por rotación durante 5 minutos a 4°C . El último lavado se realizó con 1 mL de TBS. Finalmente, se centrifugó a $960 \times g$ durante 3 minutos. Las bolas magnéticas se resuspendieron en 100 μL de solución de elución. Para eluir el complejo ADN-bolas se incubó a 65°C durante 15 minutos y se centrifugó a $20800 \times g$ por 1 minuto. El sobrenadante se guardó y se repitió el proceso de elución otra vez. El sobrenadante total, junto con las muestras control, se incubó con 0.3 M de NaCl y 3 volúmenes de solución X-link durante 16 h a 65°C .

Finalmente se procedió a la precipitación del ADN. Las eluciones se incubaron con 0.2 $\mu\text{g}/\mu\text{L}$ de enzima ARNasa A (Thermo) durante 1 h a 37°C . Posteriormente se incubó con 0.2 $\mu\text{g}/\mu\text{L}$ de Proteinasa K (Sigma) durante 2 h a 56°C . Para la precipitación, se añadió un volumen de fenol:cloroformo:isoamílico (25:24:1) y se centrifugó a $20800 \times g$ durante 5 minutos. Se añadió 50

µg de glicogen, 2 volúmenes de etanol frío, se incubó por un mínimo de 20 minutos a -20°C y se centrifugó a 20800 x g durante 10 minutos a 4°C. Finalmente se lavó el pellet con 1 mL de etanol 70% y tras la precipitación se resuspendió en 20 µl de solución TE.

El ADN precipitado se utilizó para la determinación con oligonucleótidos específicos (Tabla 7) de las regiones inmunoprecipitadas por qPCR. El ADN precipitado se normalizó respecto a su muestra control.

2.8. Obtención y análisis de ARN.

2.8.1. Extracción.

El ARN total se extrajo de células cultivadas in vitro y de muestras de tumores xenoinjertados de ratón utilizando TRIzol (Invitrogen) siguiendo las instrucciones del fabricante. Las células recogidas se lavaron en PBS frío y se centrifugaron a 240 x g a 4°C durante 5 min. El pellet de células se resuspendió en la cantidad de TRIzol recomendada por el fabricante. Para muestras de tumores xenoinjertados de ratón, la homogenización se realizó en TRIzol, según recomendación del fabricante, mediante disgregación mecánica con Kinetics Polytron Stand (Fisher Scientific). El ARN obtenido se cuantificó y se determinó su grado de pureza con un espectofotómetro Nanodrop 1000 (Thermo Scientific).

2.8.2. Retrotranscripción (RT).

El ARN total obtenido de células, de tejidos humanos o tumores se retrotranscribió a ADNc para su posterior análisis utilizando el kit GoScript™ Reverse Transcriptase (Promega) siguiendo las recomendaciones del fabricante en el termociclador T100 Thermal cycler (BioRad). Se utilizaron 500 ng de ARN total y las condiciones para la RT fueron: 5 minutos a 25°C; 120 minutos a 42°C y 15 minutos a 70°C. El ARN de tejidos humanos fue cedido por la Dra. Begoña Aguado.

2.8.3. PCR.

El producto de ADNc obtenido en la RT-PCR o el ADN obtenido de células se amplificó utilizando oligonucleótidos específicos para cada gen (Tabla 5-6) utilizando el kit GoTaq Flexi DNA Polymerase (Promega) en el termociclador T100 Thermal cycler (BioRad). Los ciclos utilizados variaron en función del experimento según las recomendaciones del fabricante. Los productos de PCR se resolvieron en gel de agarosa 1% al que se añadió GelRed (BioNova) sumergido en solución TAE 1x. Se utilizó una escalera de 1 kb (GeneRuler, Thermo Fisher) como marcadores de tamaño de ADN. Los productos de PCR se separaron mediante electroforesis y se visualizaron en un transiluminador de luz ultravioleta Uvidoc System (UVITEC).

2.8.4. PCR cuantitativa (qPCR).

El producto de ADNc obtenido en la RT-PCR se amplificó utilizando oligonucleótidos específicos para cada gen (Tabla 6) utilizando el kit GoTaq qPCR Master Mix (Promega) y siguiendo las instrucciones del fabricante en el termociclador ABI PRISM 7900HT (Applied Biosystems). Los ciclos utilizados fueron: 5 minutos a 95°C; 40 ciclos de 15 segundos a 95°C, 30 segundos a 60°C y 15 segundos a 72°C; 1 ciclo de 5 minutos a 95°C. Los niveles relativos de ARNm se obtuvieron y calcularon a partir de la señal de amplificación por encima de la línea umbral (Ct), mediante el programa ABI PRISM 7900HT SDS v2.4 Software (Applied Biosystems). La ecuación utilizada para calcular la cantidad relativa se define como: $2^{-[(Ct(\text{gen estudio}) - Ct(\text{gen control})) \text{ muestra estudio} - (Ct(\text{gen estudio}) - Ct(\text{gen control})) \text{ muestra control}]}$. Cada reacción se realizó por triplicado y los valores fueron normalizados a la expresión del gen control *GAPDH*.

2.9. Obtención y análisis de proteínas.

2.9.1. Extracción.

Las células recogidas se lavaron en PBS frío y se centrifugaron a 240 x g a 4°C durante 5 min. El pellet de células se resuspendió en solución de lisis de proteínas (100 – 500 µL) suplementado con inhibidores de proteasas (Compleat-EDTA-free, Roche). Tras la incubación a 4°C durante 30 minutos, los lisados celulares se centrifugaron a 12.000 x g a 4°C durante 15 min. El sobrenadante se guardó y se determinó la concentración de proteínas mediante BCA Protein Assay Kit (Pierce Biotechnology) siguiendo las instrucciones del fabricante. La absorbancia se midió utilizando un lector de placas Bio Kenetics Reader EL340 (Bio-Tek Instrument). La cuantificación de proteínas se determinó gracias al uso de una curva patrón de proteína BSA de concentraciones conocidas.

2.9.2. Análisis por electroforesis e inmunodetección (Western-blot).

Las proteínas obtenidas del lisado celular (10 – 20 µg) se mezclaron con tampón de carga a concentración final 1X, se incubaron a 100°C durante 5 minutos y se resolvieron mediante electroforesis por separación por peso molecular en geles de poliacrilamida en condiciones desnaturizantes al 10% (SDS-PAGE). La electroforesis se realizó en cubetas específicas para esta técnica (BioRad) aplicando un voltaje constante de 100 V en solución de electroforesis hasta alcanzar la resolución deseada. Posteriormente, las proteínas se transfirieron a una membrana de nitrocelulosa de 0.2 µm de poro (BioRad) utilizando el sistema de transferencia húmedo (BioRad) durante 1 h y 30 minutos a un voltaje constante de 100 V en solución de transferencia. El bloqueo de la membrana se realizó durante 1 hora a temperatura ambiente en 5% BSA disuelto en TBS-T. Tras este proceso, se incubó la membrana con anticuerpos específicos para cada proteína (Tabla 4) diluido en TBS-T 3% BSA durante una noche a 4°C. Las membranas se lavaron 3 veces con TBS-T a temperatura ambiente y se

incubó con el anticuerpo secundario correspondiente conjugado a peroxidasa durante 1 h en TBS-T 3% BSA a temperatura ambiente. Tras 3 lavados con TBS-T y un último con TBS, la señal quimioluminiscente se detectó utilizando el kit SuperSignal West Dura Substrate (Pierce Biotechnology) siguiendo las instrucciones de fabricante, permitiendo reaccionar la peroxidasa y potenciando la señal. Finalmente se visualizaron las proteínas detectadas tras la exposición y revelado en películas autorradiográficas (AGFA). Las bandas de proteínas fueron densitometradas mediante el programa informático ImageJ (NIH).

2.9.3. Inmunoprecipitación (IP).

Aproximadamente 1×10^7 células se utilizaron para cada condición. El lisado celular se llevó a cabo de la manera ya descrita utilizando 500 μ L de solución IP. Tras la cuantificación, 1 mg de proteína total se incubó con proteína A/G (Thermo Scientific) siguiendo las instrucciones del fabricante durante 2 h en rotación a 4°C para eliminar uniones inespecíficas. Se centrifugó a 20800 x g durante 1 minuto a 4°C y el sobrenadante se incubó con el anticuerpo α -FLAG (Sigma) a concentración 1:50 en rotación durante 12 h a 4°C. Seguidamente se incubó con proteína A/G en rotación durante 12 h a 4°C. Posteriormente, se lavaron las muestras 3 veces en solución IP centrifugando a 960 x g durante 1 minuto a 4°C. Por último, la proteína A/G conjugada con el anticuerpo y las proteínas inmunodetectadas se incubaron a 100°C durante 5 minutos en solución de carga a concentración final 1X para desacoplar las proteínas del complejo proteína A/G. El inmunoprecipitado se resolvió por Western-blot utilizando los anticuerpos específicos para la detección de las proteínas deseadas.

2.9.4. Interactoma.

El análisis de las proteínas de interacción de TCFL5 y CHA se llevó a cabo por el Servicio de Proteómica del Centro de Biología Molecular Severo Ochoa (CBMSO, Madrid, España) los cuales pertenecen al Instituto Nacional de Proteómica (ProteoRed). Brevemente, el producto de la inmunoprecipitación se resolvió por electroforesis en un gel concentrador y se realizó una digestión triptica. Las muestras obtenidas se analizaron por LC-MS/MS en un LTQ-ORBITRAP-VELO-PRO (Thermo Scientific). La identificación de proteínas se realizó con el software Proteome Discover 1.4 (Thermo Scientific) usando como motor de búsqueda SEQUEST y la base de datos de humanos de Uniprot.

2.9.5. Inmunofluorescencia.

Para la inmunofluorescencia se utilizaron tanto mESCs como embriones de ratón en diferentes estadios del periodo de preimplantación. Las mESCs se sembraron en placas de 24 pocillos sobre cristales de 12 mm de diámetro (Thermo Scientific) esterilizados y previamente colocados en la base de cada pocillo y tratados con gelatina 0.2 %. Tanto las mECSs como los embriones se

fijaron en 4% de PFA libre de metanol durante 10 minutos a temperatura ambiente. Tras dos lavados con PBS, las muestras se permeabilizaron en solución de permeabilización durante 30 minutos a temperatura ambiente y posteriormente se bloquearon en solución de bloqueo durante 1 hora a temperatura ambiente. Seguidamente se realizó la incubación con el anticuerpo específico (Tabla 4) diluido en solución de tinción durante la noche a 4°C. Tras 3 lavados con PBS-T a temperatura ambiente, se incubaron las muestras con el correspondiente anticuerpo secundario conjugado con fluoróforo durante 1 h a 4°C. Se realizaron 3 lavados con PBS-T y un último con PBS. Finalmente, las preparaciones se montaron sobre portaobjetos utilizando medio de montaje VectaShield con Dapi (Vector Laboratories) siguiendo las recomendaciones del fabricante. La visualización de la fluorescencia se llevó a cabo utilizando un microscopio confocal LSM510 acoplado a una cámara Axiolmager M1 (Zeiss). El procesamiento de la imagen se realizó en el programa ImageJ (NIH).

2.9.6. Análisis por citometría de flujo.

Para el análisis de ciclo celular, las células se crecieron en placas de cultivo de 6 pocillos y se trataron durante 24 h con 10 mM de KarioMax (Thermo Scientific). Se recolectaron las células por tripsinización, se lavaron 3 veces en PBS frío y se fijaron añadiendo suavemente 1 mL de etanol 70% frío. Las muestras fijadas se mantuvieron a -20°C por un periodo mínimo de 18 h. Tras este proceso, las muestras se lavaron 3 veces con PBS y se centrifugaron a 240 x g durante 5 minutos a 4°C. Previo al análisis, las muestras se marcaron con tampón de ciclo PI/RNase Staining Buffer (BD Bioscience) durante 30 minutos en oscuridad a temperatura ambiente. El análisis por citometría se realizó en un citómetro de flujo FACSCalibur (BD Bioscience) y el programa FlowJo (FlowJo, LLC, EEUU).

Para el análisis de la actividad Aldehído Deshidrogenasa se utilizó el kit Aldefluor (StemCell Technologies) siguiendo las instrucciones del fabricante. El análisis se realizó en un citómetro de flujo FACSCalibur (BD Bioscience) y el programa FlowJo (FlowJo, LLC, EEUU).

Para el análisis de la expresión de los marcadores de CSCs, las células recolectadas se lavaron en PBS 1% BSA frío. Las células se incubaron con los anticuerpos específicos (Tabla 4) durante 20 minutos en oscuridad a 4°C. Se lavaron las células con PBS 1% BSA frío y se incubó con los respectivos anticuerpos secundarios acoplados a fluorocromos durante 20 minutos en oscuridad a 4°C. Se lavaron las células, se resuspendieron en PBS 1% BSA y se analizaron por citometría. El análisis se realizó en un citómetro de flujo FACSCanto A (BD Bioscience) y el programa FlowJo (FlowJo, LLC, EEUU).

2.10. Ensayos de luminometría.

Las células se sembraron en placas de 24 pocillos a una densidad del 70%. Las células fueron transfectadas con el plásmido de actividad luciferasa deseado para cada experimento junto con un

plásmido de expresión constitutiva de la luciferasa *Renilla*, como control de transfección. Para la transfección se utilizó Metafectene Pro (Biontex) siguiendo las instrucciones del fabricante. Tras 48 horas, se realizó el análisis de la actividad luciferasa utilizando el kit *Dual-Luciferase Reporter Assay System* (Promega) siguiendo las instrucciones del fabricante. Se cuantificó la luz emitida utilizando el lector FLUOstar OPTIMA (BMG LabTech).

2.11. Proliferación.

2.11.1. Contaje celular.

Las células se sembraron en placas de 6 pocillos a una densidad de 1×10^4 células por pocillo. Se utilizó un pocillo por cada día de contaje. Durante 12 días, se levantaron las células cada 2 o 3 días por tripsinización y se contó el número de células utilizando una cámara de contaje celular Neubauer.

2.11.2. Ensayo colorimétrico MTT.

Las células se sembraron en placas de 96 pocillos a una densidad de 5×10^3 células por pocillo durante 24 y 48h. Se añadió el reactivo 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) al 10% (v/v) al medio a tiempo final del experimento. Tras 4 h de incubación, se eliminó el sobrenadante y se añadió 150 μ L de DMSO (Sigma Aldrich) e isopropanol (Sigma) en proporciones 1:1 junto con ácido acético 0.3% (Sigma). Se agitó durante 2 minutos y se leyeron las absorbancias a 495 nm en un lector multiplaca iMark (BioRad).

2.12. Formación de colonias.

Las células se sembraron en placas de 6 pocillos con una densidad de 500 células por pocillo. El medio de cultivo fue renovado cada 2 días durante las 2 semanas de duró el experimento. A día final, las células se lavaron 2 veces con PBS y se fijaron en 4% de PFA (Sigma) durante 20 minutos a temperatura ambiente. Las células fijadas se lavaron 3 veces con PBS. A continuación, se tiñeron con solución Cristal Violeta durante 30 minutos a temperatura ambiente. Posteriormente se lavaron las células por inmersión en abundante agua y se dejaron secar. Para el contaje del número de colonias, se dividió la placa en dos y se contaron manualmente el número de estas.

2.13. Migración por cerrado de herida.

Las líneas celulares se cultivaron en placas de 6 pocillos a un 80% de confluencia. Una vez las células se encontraron adheridas, se realizó una herida en la monocapa de células de 0.3-0.4 mm de ancho. Posteriormente, se lavaron las células 2 veces con PBS estéril para retirar las células

desprendidas. Finalmente, se añadió medio MEM completo con 0.4% FBS que fue reemplazado cada día. Cada 24 horas durante 3 días, se tomaron imágenes de la herida en un microscopio Leica DM IL (Leica Microsystems). Se cuantificó el área de la herida utilizando el programa ImageJ (NIH). Se normalizó el área con respecto al día inicial de cada condición representado el porcentaje de herida cerrada.

2.14. Tamaño y número de MCTS.

Las líneas celulares se cultivaron en placas de 96 pocillos de baja adhesión celular para la formación de MCTS. Tras 7 días de cultivo, se tomaron fotografías en un microscopio Leica DM IL (Leica Microsystems). Los MCTS fueron analizados en el programa ImageJ (NIH) y se tomaron como válidos aquellos con un área mayor a 250 μm . El tamaño se calculó normalizando el área de los MCTS con respecto al área media de la condición control.

2.15. Generación de tumores xenoinjertados.

Se utilizaron ratones hembra inmunodeprimidos Swiss Nude (Crl:NU(Ico)-Foxn1nu) (Charles River Laboratory). Para el caso de la línea celular HT29-D6-luc, el experimento se realizó por duplicado en grupos de 5 ratones inyectando de manera subcutánea en el flanco posterior del animal 1×10^6 células de cada línea sobreexpresada. En el caso de la línea celular HCT116, se utilizaron grupos de 5 ratones inyectando de manera subcutánea en ambos flancos posteriores del animal 2×10^6 células de las líneas sobreexpresadas y deficientes. El crecimiento tumoral fue seguido mediante medición del volumen del tumor durante 7-8 semanas consecutivas hasta finalizar el experimento. El volumen del tumor se midió utilizando un calibre milimétrico y se estimó siguiendo la ecuación $\{\text{ancho}\} \times \{\text{alto}\} \times \{\text{largo}\}$. Además, en el caso de las HT29-D6-luc, se siguió la luminiscencia a lo largo del experimento inyectando intraperitonealmente 150 μg de luciferina por cada gramo del peso del ratón. Tras 10 minutos desde la inoculación para la correcta distribución por el organismo, se anestesiaron los animales con isoflurano (Forane) por inhalación mediante un equipo de anestesia gaseoso. La emisión de bioluminiscencia fue recogida por el equipo IVIS Lumina II (Caliper Life Science) y se cuantificó con el programa Living Image Software (Perkin Elmer) del RedLab Biolumlab de la Comunidad de Madrid (CBMSO, CSIC-UAM). A día final de experimento, los animales fueron sacrificados y los tumores xenoinjertados extraídos para su posterior análisis.

2.16. Ética.

Los animales de experimentación, su mantenimiento y los experimentos cumplieron con la legislación vigente Nacional Española y de la Unión Europea (PROEX 21/14). El trabajo fue supervisado y avalado por el Comité Ético de Bienestar Animal del CBMSO. La Lda. Beatriz Barrocal, que cuenta con el certificado de experimentación animal de categoría B y C, fue la veterinaria

responsable de los animales supervisando, aconsejando y dirigiendo el manejo de estos. A su vez, el doctorando cuenta con el certificado de experimentación animal de categoría B. Todos los animales fueron sacrificados mediante inhalación de CO₂ a la finalización del experimento. En caso de pérdida de calidad de vida por parte de los animales por causa directa o indirecta de la experimentación, los animales fueron sacrificados.

2.17. Estadística.

Los análisis estadísticos se realizaron con el programa GraphPad Prism v6.0 (GraphPad Software, LLC). El test de t-Student fue utilizado para comparar las medias de datos cuantitativos. En el caso de gran cantidad de datos, se identificaron los *outliers* y se eliminaron del análisis estadístico. Todos los experimentos se representaron como la media \pm SEM (error estándar de la media). Los resultados se consideraron estadísticamente significativos cuando el valor del p-valor fue igual o menor a 0.05 representándose como * $p < 0.05$, ** $p < 0.01$ y *** $p < 0.001$.

Resultados

Resultados

1. Descripción y expresión tisular en condiciones fisiológicas de las isoformas del factor de transcripción *TCFL5* en humano y de ratón.

A pesar de que *TCFL5* fue descrito en el año 1998, dada la escasa literatura acerca de este factor de transcripción, su función es poco conocida. La mayor parte de estudios publicados y estudios en microarrays no discriminan entre la expresión de las distintas isoformas del gen y, por tanto, no se puede atribuir una función específica a cada una de ellas. Debido al potencial interés del gen en distintos modelos de desarrollo y cáncer, nos propusimos describir las isoformas que presenta el gen *TCFL5*, en tejidos humanos y de ratón, primero según diversas bases de datos de microarrays y secuenciación de ARN (RNA-Seq) y posteriormente de forma experimental.

1.1. *TCFL5* se expresa de forma ubicua en humano principalmente en testículo y cerebro.

En primer lugar, quisimos analizar la expresión de *TCFL5* en los distintos órganos humanos utilizando datos de RNA-Seq y microarrays depositados en diferentes bases de datos. Obtuvimos datos de RNA-Seq procedentes de dos proyectos independientes: “*Genotype Tissue Expression project*” (GTEx) [127] y “*Human Protein Atlas project*” (HPA) [93] (Figura 7A). En ambas colecciones se puede observar que el órgano con una mayor expresión de *TCFL5* a nivel de ARN es el testículo. El cerebro es el segundo órgano con mayor expresión mientras que el resto de tejidos presenta una expresión mucho más baja, aunque detectable en todos ellos. Para una mayor consistencia en los datos, decidimos comprobar su expresión utilizando datos procedentes de microarrays. Se unificaron varios experimentos que utilizaron el array de expresión “*Human Genome U133 Plus 2.0 Array*”. Este array presenta dos sondas distintas para la detección de *TCFL5*: 204849_at y 235694_at. La sonda 204849_at mapea en el último exón del gen por lo que es capaz de reconocer las isoformas descritas *TCFL5* y *CHA* mientras que la sonda 235694_at lo hace en una región entre el último y penúltimo exón, lo que no se corresponde con ninguna isoforma descrita previamente en la literatura (Figura 7B). En estos arrays la mayor expresión de *TCFL5* se encuentra en testículo, seguido de tejidos de cerebro y, por último, en el resto de tejidos concordando con los datos de RNA-Seq (Figura 7C). Además, la expresión de la sonda 204849_at fue mayor que la de la sonda 235694_at en la mayoría de órganos. Sin embargo, en ciertos tejidos como son pituitaria, nódulos linfoides, ventrículo del corazón, glándula tiroidea, colon y pulmón, la expresión de estas sondas se igualó e incluso se llegó a invertir el patrón de expresión en oocitos. Estos datos indican que puede existir una isoforma no descrita que se expresa en todos los órganos, aunque en menor cantidad.

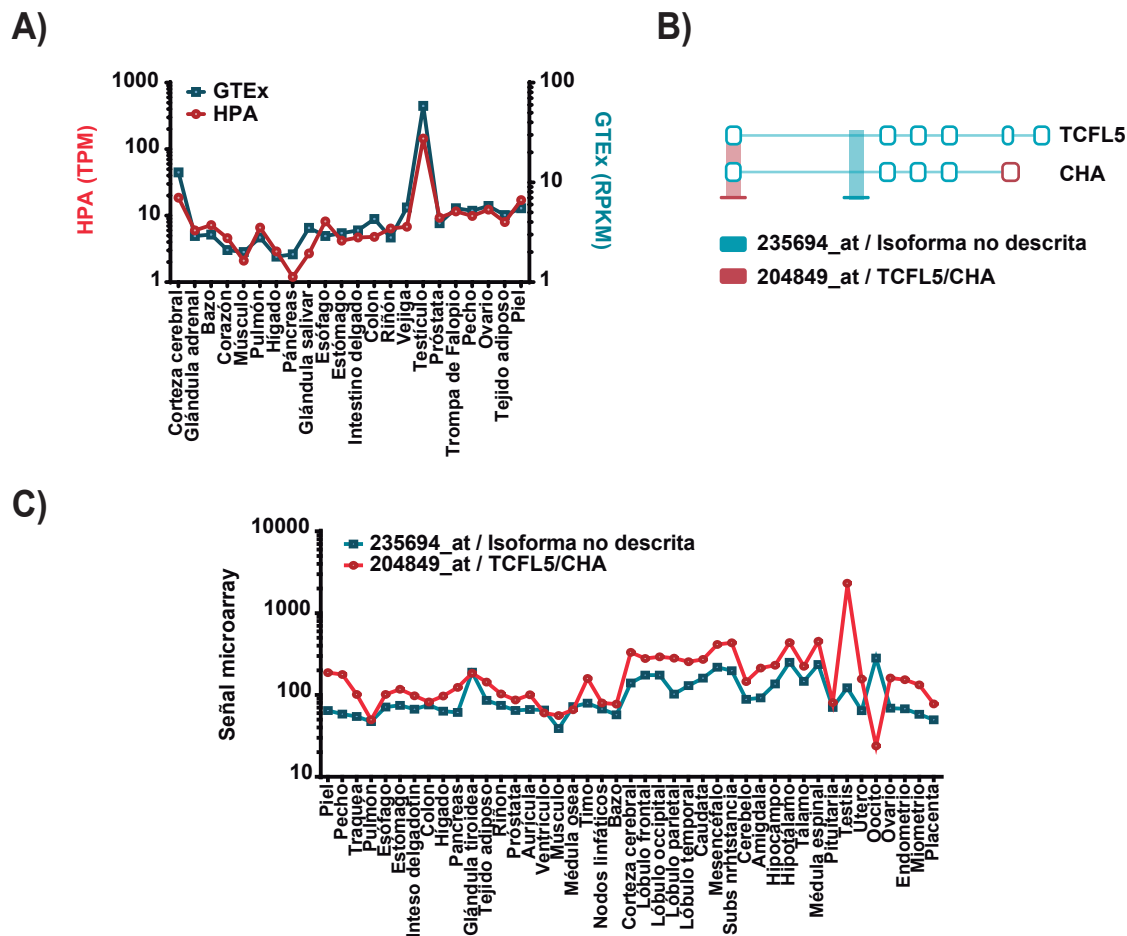


Figura 7. *TCFL5* se expresa en todos los tejidos humanos. **A)** Análisis de la expresión de *TCFL5* por RNA-Seq de los proyectos *Human Protein Atlas project* (HPA; rojo) y *Genotype Tissue Expression project* (GTEx; azul). **B)** Esquema de las regiones de reconocimiento de las sondas 204849_at y 235694_at utilizadas en el microarray *Human Genome U133 Plus 2.0 Array*. **C)** Expresión de las sondas de *TCFL5* 204849_at (rojo) y 235694_at (azul) en distintos órganos humanos con el microarray *Human Genome U133 Plus 2.0 Array*.

1.2. *TCFL5* presenta varias isoformas en humano.

Ante la aparición de una posible nueva isoforma, realizamos una búsqueda de todas las secuencias, descritas y predichas para *TCFL5* anotadas en base de datos (NCBI, Ensembl, Vega, ECGene y Uniprot). Tras su alineamiento, generamos las variantes consenso del gen (Tabla 12). Describimos 10 exones diferentes a los que llamaremos: E1, E2, E2b, E3, E4, E4b, E5, E6, E7 y E8. Estos exones se combinan de distinta manera para dar lugar a un total de 6 isoformas a las que hemos nombrado de la siguiente manera: *TCFL5_E8* o isoforma canónica *TCFL5* (E1, E2, E3, E4, E5, E8), *TCFL5_E8b* (E1, E2, E3, E4b, E5, E8), *CHA* (E2b, E3, E4, E5, E8), *TCFL5_E6* (E1, E2, E3, E4, E5, E6), *TCFL5_E6b* (E1, E2, E3, E4b, E5, E6) y *TCFL5_E7* (E1, E2, E3, E4, E5, E7) (Figura 8A). Los exones E2, E3, E4 y E5 componen un núcleo de exones comunes a todas las isoformas mientras que el resto de exones son específicos para cada una de ellas. Así, las distintas isoformas se diferencian en el primer y/o último exón que utilizan. Existen 3 exones alternativos para el último exón: E8, E7 y E6. La expresión del exón E8 permite distinguir *TCFL5_E8*, *TCFL5_E8b* y *CHA* del resto de

isoformas. El exón E6 es específico para las isoformas *TCFL5_E6* y *TCFL5_E6b* y el exón E7 lo es para la isoforma *TCFL5_E7*. Con respecto al exón inicial, existen dos exones alternativos: E1 y E2b. El exón E1 es el más grande y representa la mitad del gen. Este exón se encuentra presente en todas las isoformas, excepto en *CHA*. El exón E2b es exclusivo de la isoforma *CHA* y comparte casi toda la secuencia con el exón E2. Sin embargo, el exón E2b presenta un inicio de traducción alternativo usando una metionina presente al inicio de este exón siendo los nucleótidos anteriores un posible 5'UTR. Por último, el exón E4b es un exón alternativo que utilizan las isoformas *TCFL5_E8b* y *TCFL5_E6b* en lugar del exón E4. Este exón carece de los 3 nucleótidos iniciales del exón E4 por lo que codifica para un aminoácido menos. No se encontró ninguna secuencia que soporte alguna isoforma que utilice el exón inicial E2b con los finales E6 y E7.

	NCBI	Ensembl	Vega	Ecgene	Uniprot
Gene ID	10732	ENSG0000101190	OTTHUMG00000032939	-	-
<i>TCFL5_E8</i>	NM_006602 AB012124.1 AF070992.1	ENST00000335351	OTTHUMT00000080079	H20C8515.6 H20C8515.3	Q9UL49-3 Q9UL49-1
<i>TCFL5_E8b</i>	XM_005260185.1	-	-	-	-
<i>TCFL5_E7</i>	XM_005260184.1 BC046933.1 H20C8515.8	ENST00000217162	-	-	Q86TP4 F8W9A4
<i>TCFL5_E6</i>	H20C8515.9	-	-	-	-
<i>TCFL5_E6b</i>	XM_005260186.1 BC065520.1	-	-	-	-
<i>CHA</i>	AJ271337.1	-	-	H20C8515.5 H20C8515.2	Q9UL49-2

Tabla 12. Tabla resumen de todas secuencias de *TCFL5* en humano depositadas en las distintas bases de datos (NCBI, Ensembl, Vega, Ecgene y Uniprot).

1.3. Las isoformas de *TCFL5* se expresan en todos los órganos humanos.

Tras el alineamiento y la descripción de las variantes consenso para *TCFL5*, quisimos estudiar la expresión de las isoformas en los diferentes órganos humanos. Obtuvimos datos de RNA-Seq de tejidos y células (Tabla S1) y calculamos la cobertura media normalizada (CMN) de cada exón para inferir la incorporación de estos a los mensajeros. Para calcular esta CMN, normalizamos con la cobertura media de la región intrónica comprendida entre los exones E4 y E5 y entre la cobertura media del gen (Figura 8B). Como era de esperar, los exones comunes E2, E3, E4 y E5 son los que mayor CMN presentaron y por tanto los más incorporados en las isoformas que se expresan en todas las células y tejidos analizados. No aparecen diferencias significativas entre las CMN de los exones E6, E7 y E8 por lo que su expresión no parece específica de los tejidos y células analizados. Sin embargo, sí existe una CMN mayor para el exón E8, seguido del E6 y por último el E7 indicando una mayor incorporación del E8, que del E6 y E7. La distribución del exón E1 no parece seguir el patrón de expresión de los exones centrales E2, E3, E4 y E5 por lo que debe de haber un inicio de transcripción en el exón E2b correspondiente con la isoforma *CHA*. Para poder diferenciar entre los dos inicios de transcripción alternativo, calculamos la CMN específica para *CHA*. Utilizamos la CMN del exón E2b que no comparte con el exón E2. Sin embargo, la CMN de esta región no fue muy

elevada por lo que no se pudo incorporar al estudio.

Una vez analizada la incorporación de los exones en las isoformas expresadas en los distintos órganos humanos, estudiamos en los datos de RNA-Seq las uniones de exones más frecuentes para verificar las isoformas previamente descritas. Se consideraron válidas para el análisis las uniones de exones soportadas por al menos dos lecturas en una misma muestra o, al menos, por dos muestras y fueron normalizadas por millón de lecturas alineadas en cada muestra. Tras este filtro, se detectaron

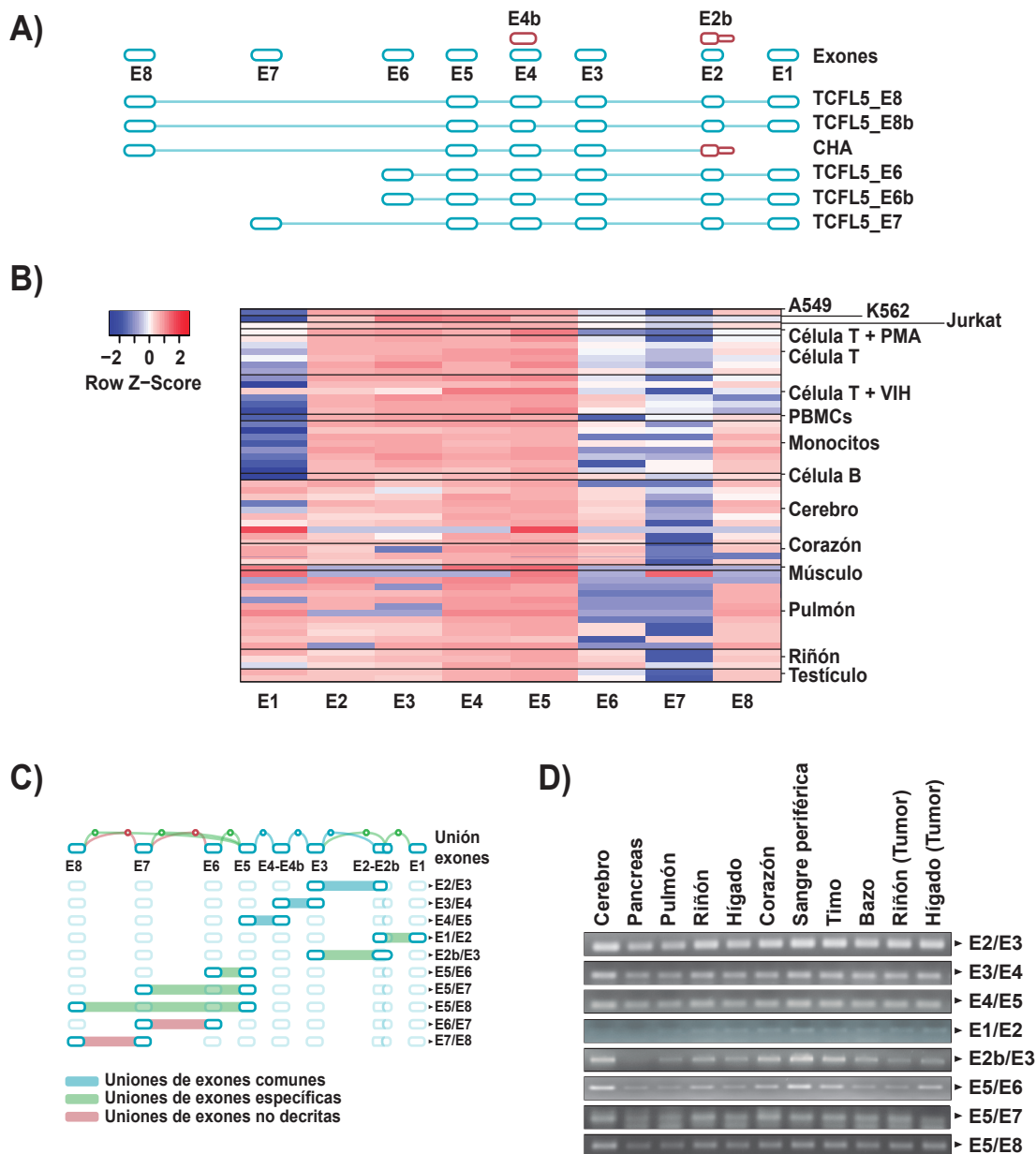


Figura 8. *TCFL5* presenta varias isoformas expresadas en los distintos tejidos humanos. **A)** Esquema consenso de los exones E1, E2, E2b, E3, E4, E4b, E5, E6, E7 y E8 y las isoformas *TCFL5_E8*, *TCFL5_E8b*, *CHA*, *TCFL5_E6*, *TCFL5_E6b* y *TCFL5_E7* a partir de las secuencias anotadas en las diferentes bases de datos. **B)** Análisis de la cobertura media normalizada (CMN) obtenida por RNA-Seq de los exones E1, E2, E3, E4, E5, E6, E7 y E8 en distintas muestras de órganos humanos. **C)** Esquema de las uniones de exones comunes (azul), específicas (verde) y no descritas (rojo) encontradas en los estudios de RNA-Seq. **D)** Amplificación por PCR de las uniones de exones E1/E2, E2b/E3, E3/E4, E4/E5, E5/E6, E5/E7 y E5/E8 en distintas muestras de órganos humanos.

un total de 51 posibles uniones de exones de las cuales solo las uniones número 3, 8, 32, 38, 41, 43, 48 y 50 fueron alta y homogéneamente expresadas en la mayoría de muestras analizadas (Figura S1). Seleccionamos como válidas las uniones de exones E2/E3, E3/E4 y E4/E5 como uniones comunes y E1/E2, E2b/E3, E5/E6, E5/E7 y E5/E8 como uniones específicas de cada isoforma (Figura 8C). En ningún caso encontramos como válidas las uniones E6/E7 y E7/E8 lo que confirma los datos del alineamiento previo.

Por último, para confirmar inequívocamente la presencia de estos exones en los distintos tejidos humanos y las uniones de exones descritas, amplificamos, por PCR, estas zonas en muestras de tejidos humanos (cerebro, páncreas, pulmón, riñón, hígado, corazón, sangre periférica, timo, bazo y tumores de riñón e hígado) (Figura 8D). Detectamos las uniones de exones comunes E2/E3, E3/E4 y E4/E5, y también la unión E5/E8, expresándose de manera homogénea en todos los tejidos. La unión de exones E1/E2 se detectó en todos los tejidos excepto en cerebro, páncreas y pulmón. La unión E2b/E3 se detectó en todos los tejidos excepto en páncreas. Las uniones de exones E5/E6 y E5/E7 presentaron diferencias de expresión entre tejidos, aunque detectable en todos ellos.

1.4. La expresión de *TCFL5* se encuentra bajo el control de diversas vías de señalización.

Los datos obtenidos confirman la existencia de al menos 4 isoformas de *TCFL5* (*TCFL5_E8* o *TCFL5*, *CHA*, *TCFL5_E6* y *TCFL5_E7*) y que todas expresan el motivo funcional bHLH expresándose en prácticamente todos los órganos humanos y en mayor medida en testículo y cerebro. No existe un exón final específico de órgano, sin embargo, la expresión de los exones E1 y E2 no se correlacionó con el resto por lo que decidimos analizar más en detalle la cobertura de estos exones. Observamos 3 escenarios diferentes con respecto a la cobertura de estos exones (Figura 9A). En células B, no se observó cobertura en el primer exón y se detectó a partir del exón E2b lo que indicaría la expresión prioritaria de la isoforma *CHA*. En testículo se observó cobertura en el exón E1, pero no en el exón E2b, indicando que no habría expresión de *CHA*. En sangre, se observó un salto en la cobertura del primer exón que no puede explicarse por errores de alineamiento puesto que aparecerían en todas las muestras e indican una expresión distinta a la estudiada para el exón E1. Estos datos indican que los exones E1 y E2 podrían tener inicios de transcripción independientes y por tanto que podrían estar controlados por dos promotores independientes. Para estudiar esta posibilidad, analizamos el promotor de *TCFL5* utilizando al proyecto *Roadmaps Epigenomics* [128]. En este proyecto se analizaron 12 marcas epigenéticas en 127 epigenomas de referencia que comprenden distintas modificaciones de histonas para determinar una predicción consenso del estado de la cromatina. Como se puede observar en la figura 9B, la región promotora de *TCFL5* comprende la región Chr20: 61.491.400-61.494.400 coincidiendo con el exón E2 hasta el 5'upstream del exón E1. Según estas predicciones, el promotor de *TCFL5* presenta una zona de *enhancer* en la región Chr20: 61.493.400-61.494.400 seguido de un promotor pausado (*poised promoter*) en la

región Chr20: 61.492.200-61.493.400 que se caracteriza por su rápida activación e inhibición, y un promotor por debajo del inicio de transcripción en la región Chr20: 61.491.400-61.492.200. Dentro de la región con el promotor pausado se encuentran los dos posibles inicios de transcripción: uno que coinciden con el inicio de la transcripción del exón E1 y otro que coincide con el inicio de la transcripción del exón E2b. Estos datos sugieren que el promotor de *TCFL5* se encuentra en un estado de rápida activación e inhibición pudiendo controlar de esta manera la expresión del exón E1 o E2b a través de dos distintos inicios de transcripción.

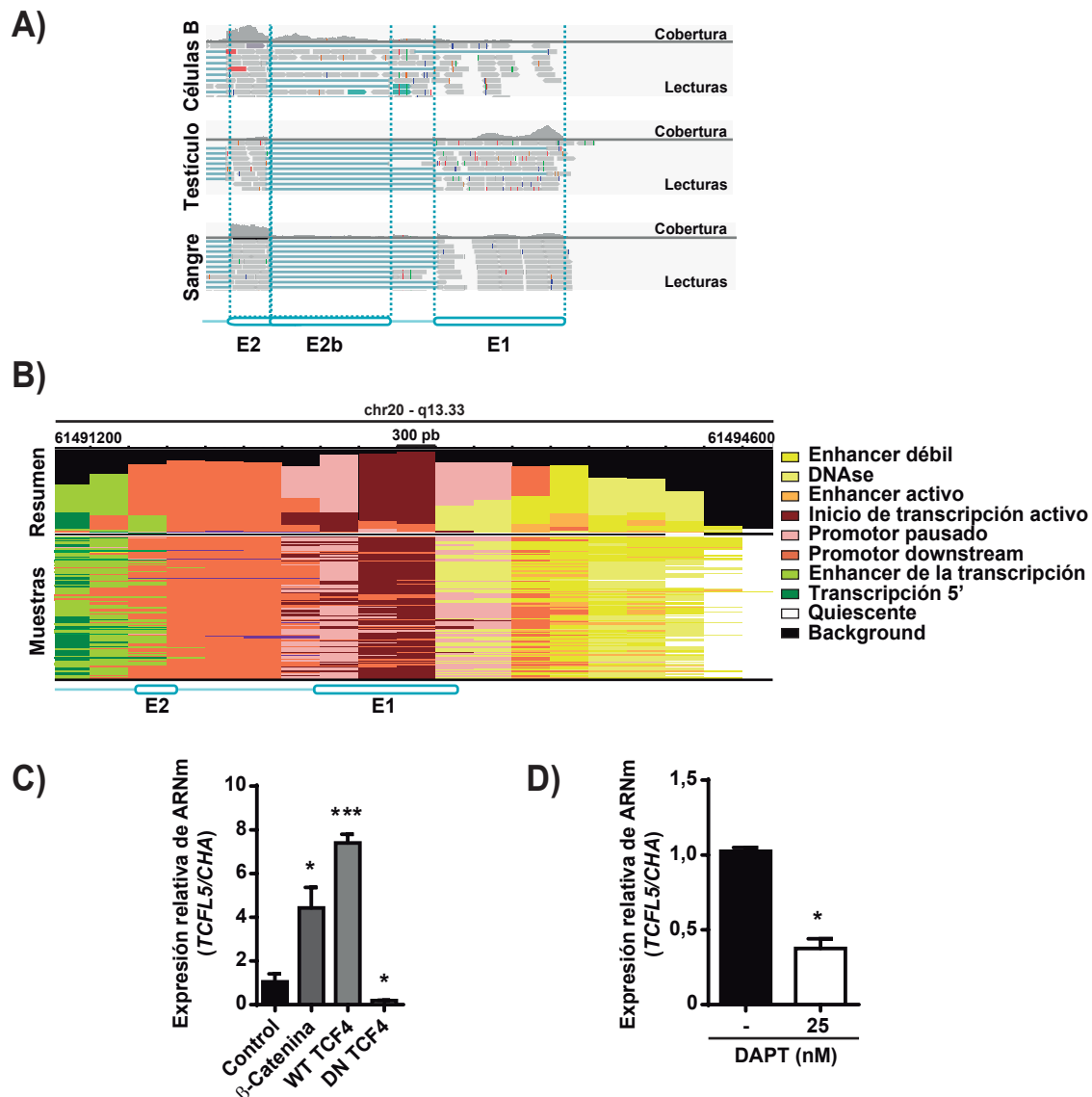


Figura 9. *TCFL5* presenta un promotor pausado y es regulado por las vías de WNT y NOTCH1. **A)** Cobertura y lecturas de RNA-Seq en el exón E1, E2 y E2b de *TCFL5* en muestras de células B, sangre y testículo humano. **B)** Esquema modelo resumen del estado de la cromatina en la región del exón E1 y exón E2 de *TCFL5* basado en 25 marcas epigenéticas (*ROADmaps Epigenomics*). **C)** Análisis por qPCR de la expresión de *TCFL5/CHA* tras la transfección con plásmido control, β -catenina, TCF4 funcional (WT-TCF4) y TCF4 dominante negativo (DN-TCF4) en células de cáncer de colon HCT116. **D)** Análisis de la expresión por qPCR de *TCFL5/CHA* tras el tratamiento con 25nM de DAPT durante 24h en células de cáncer de colon HT29. En las gráficas se representa la media \pm SEM de tres experimentos independientes y su estadística respecto al control (*p<0,05; ***p<0,001).

Tras la caracterización del promotor, estudiamos la regulación transcripcional de éste. Utilizamos datos del proyecto *ENCODE* [129] donde se analizaron CHIP-Seq de 161 factores de transcripción en 91 células distintas. Encontramos 72 factores de transcripción distintos que se unen al promotor de *TCFL5* (Tabla S2). De ellos, 53 factores de transcripción se unen 5'upstream del primer inicio de transcripción y 19 entre el primer y segundo inicio de transcripción. Además, analizamos experimentos de CHIP-Seq independientes donde observamos que otros factores de transcripción como son TCF4 (región -391), NANOG (-358 y -916) y PPARg (-1023) también se unen al promotor de *TCFL5*, entre otros. Estos datos sugieren dos posibles vías de señalización que podrían regular la expresión de *TCFL5* como son la vía de WNT y la vía de NOTCH1, previamente descrita en la literatura [104]. Para comprobar la regulación de *TCFL5* por la vía de WNT, se transfectaron en la línea celular HCT116 vectores para la sobreexpresión de β -catenina, TCF4 funcional (WT-TCF4) y un dominante negativo de este (DN-TCF4) (Figura 9C). La sobreexpresión de β -catenina y de WT-TCF4 aumentaron la expresión del ARNm de *TCFL5/CHA* mientras que la de DN-TCF4 la redujo indicando una posible regulación de estas isoformas por esta vía. Del mismo modo, comprobamos la regulación transcripcional de *TCFL5* por la vía de NOTCH1. Tratamos la línea celular HT29 con DAPT, que inhibe el corte proteolítico de NOTCH1 por la γ -secretasa y la liberación del dominio funcional ICN de este (Figura 9D). Se observó que la inhibición de la vía de NOTCH produjo una disminución de la expresión de *TCFL5/CHA* sugiriendo una regulación por parte de esta ruta de señalización.

1.5. Expresión de *Tcfl5* en ratón.

Si en humanos los datos publicados acerca de la expresión de *TCFL5* son escasos, en ratón son prácticamente inexistentes. Por ello, nos propusimos describir las isoformas de *Tcfl5* presentes en ratón. Primero, estudiamos su expresión analizando la información depositada en bases de datos. En este caso obtuvimos los datos del proyecto “*FANTOM5*” [130] (Figura 10A). En ratón, el único órgano donde aparece expresión de *Tcfl5* es en testículo. El cerebro, el segundo órgano en expresión en humanos, no presenta ninguna expresión. En el resto de órganos tampoco se detectó. Para continuar con el análisis y aumentar el estudio de expresión de *Tcfl5*, obtuvimos de bases de datos nuevos datos de RNA-Seq (Tabla S3). Observamos que, entre los órganos estudiados, la expresión fue exclusiva de testículo (Figura 10B).

1.6. *Tcfl5* presenta varias isoformas en ratón.

Al igual que hicimos en humanos, para describir las posibles isoformas de *Tcfl5* en ratón realizamos una búsqueda en bases de datos de las secuencias descritas y predichas para generar las variantes consenso (Tabla 13). Tras su alineamiento, describimos 8 exones a los que llamaremos: e1, e2, e3, e4, e4b, e4l, e5 y e6 que podrían dar lugar a 5 isoformas a las que nos referiremos como:

Tcfl5_e6 o *Tcfl5* consenso (e1, e2, e3, e4, e5 y e6), *Tcfl5_e6b* (e1, e2, e3, e4b, e5, e6), *Tcfl5_e6Δ5* (e1, e2, e3, e4 y e6), *Tcfl5_e4l* (e1, e2, e3 y e4l) y *Tcfl5_e5* (e1, e3, e4 y e5) (Figura 10C). A diferencia de humano, existe un núcleo de exones centrales que se compone de los exones e1, e2 y e3 y solo se diferencian en sus exones finales. Las isoformas *Tcfl5_e6* y *Tcfl5_e6b* se diferencian en la presencia del exón e4 y e4b, respectivamente. Este exón e4b carece de los 3 primeros nucleótidos del exón e4 y por tanto codifica para un aminoácido menos. La isoforma *Tcfl5_e6Δ5* es igual que la *Tcfl5_e6*, pero carece del exón e5 por lo que no expresaría un dominio bHLH funcional. La isoforma *Tcfl5_e4l* termina en el exón e4l y no expresa el dominio bHLH funcional mientras que la isoforma *Tcfl5_e5* carece del exón e2 y utiliza el exón final e5. En el caso de ratón, no encontramos ninguna secuencia que soporte la existencia de una isoforma igual que la denominada *CHA* humana. Para comprobar estos datos obtuvimos muestras de testículo y amplificamos por PCR las isoformas completas (Figura 10D). Utilizando oligonucleótidos específicos para la región e1/e6 obtuvimos dos amplicones: uno de mayor tamaño correspondiente a las isoformas *Tcfl5_e6* y *Tcfl5_e6b* y otro de menor tamaño debido a la falta del exón e5 correspondiente a la isoformas *Tcfl5_e6Δ5*. Utilizando oligonucleótidos específicos para la región e3/e5 solo se amplificaron las isoformas *Tcfl5_e6* y *Tcfl5_e6b*.

	NCBI	Ensembl	Vega	Uniprot
Gene ID	277353	ENSMUST00000131881	OTTMUSG00000016345	-
<i>Tcfl5_e6</i>	NM_178254.3 BC108399.1	ENSMUST0000037877	OTTMUST00000039295	Q32NY8
<i>Tcfl5_e6Δ5</i>	XM_006500653.1	-	-	-
<i>Tcfl5_e6b</i>	AK132721.1	ENSMUST00000131881	-	Q3V133
<i>Tcfl5_e4l</i>	AY234363.1 XM_006500652.1	-	-	Q810E9
<i>Tcfl5_e5</i>		ENSMUST00000161425	OTTMUST00000086954	ESQ5K1

Tabla 13. Tabla resumen de todas secuencias de *Tcfl5* en ratón depositadas en las distintas bases de datos (NCBI, Ensembl, Vega y Uniprot).

Tras el alineamiento y caracterización de las variantes consenso para *Tcfl5*, estudiamos las uniones de exones más frecuentes. Para ello analizamos los datos de RNA-Seq de tejidos de ratón obtenidos previamente y se realizó el análisis de igual manera que en humano. Se detectaron un total de 8 posibles uniones de exones (Figura S2). De esta forma, seleccionamos como válidas las uniones de exones e1/e2, e2/e3 y e3/e4 como uniones comunes y e1/e3, e3/e4l, e4/e5, e4/e6 y e5/e6 como uniones específicas (Figura 10E). Se analizaron muestras de testículo para comprobar las uniones de exones descritas mediante PCR (Figura 10F). Se encontraron expresadas las uniones de exones e1/e2, e2/e3, e3/e4, e3/e4l y e5/e6. Sin embargo, la unión e4l/e5 no se expresó como era de esperar puesto que no se detectó ninguna isoforma que presente esta unión de exones.

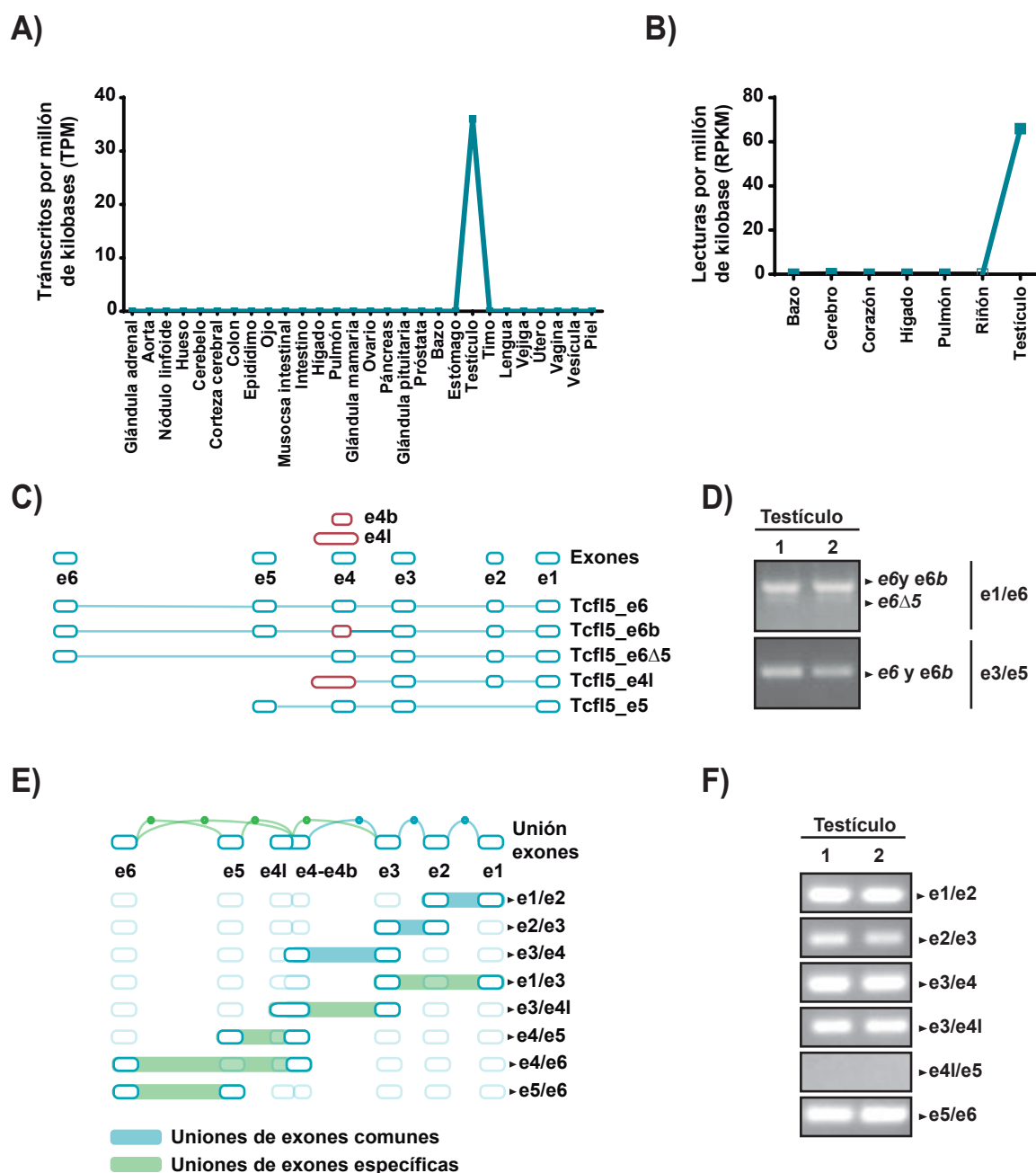


Figura 10. La expresión de *Tcf15* es específica de testículo y presenta varias isoformas en ratón. **A)** Expresión en ratón de *Tcf15* en microarrays del proyecto FANTOM5. **B)** Análisis de la expresión por RNA-Seq de *Tcf15* en distintas muestras de órganos de ratón. **C)** Esquema consenso de los exones e1, e2, e3, e4, e4b, e4l, e5 y e6 y las isoformas *Tcf15_e6*, *Tcf15_e6b*, *Tcf15_e6Δ5*, *Tcf15_e4l* y *Tcf15_e5* a partir de las secuencias anotadas en diferentes bases de datos. **D)** Amplificación por PCR de las isoformas completas *Tcf15_e6*, *Tcf15_e6b*, *Tcf15_e6Δ5* en muestras de testículo de ratón. **E)** Esquema de las uniones de exones comunes (azul) y específicas (verde) descritas por RNA-Seq. **F)** Amplificación por PCR de las uniones de exones e1/e2, e2/e3, e3/e4, e3/e4b, e4b/e5 y e5/e6 en muestras de testículos de ratón.

2. Análisis de la expresión en cáncer de colon y caracterización de la función específica de las isoformas TCFL5 y CHA en procesos tumorales.

Como se ha mencionado anteriormente, varios son los trabajos que sugieren una función de TCFL5 en cáncer de colon en distintos aspectos como proliferación, viabilidad celular o progresión, aunque sin llegar a profundizar en ellos. Sin embargo, la expresión basal presente en este tejido es

baja (Figuras 7A y 7C) y su función fisiológica es totalmente desconocida. En el siguiente apartado del presente trabajo, nos propusimos estudiar la expresión de las distintas isoformas de TCFL5 en este tipo de cáncer y el papel que juegan cada una de ellas en distintos procesos tumorales.

2.1. La expresión de *TCFL5* se encuentra inducida en cáncer de colon.

En primer lugar analizamos la expresión de *TCFL5* en distintas colecciones de muestras humanas de cáncer colorrectal utilizando para ello la base de datos *OncoMine* [131]. Obtuvimos datos de dos colecciones independientes de microarrays: *Skrzypcak Colorectal 2* [132] y *Hong Colorectal* [133]. Ambas colecciones utilizaron el microarray de expresión *Human Genome U133 Plus 2.0 Array* con las sondas para la detección de *TCFL5* 204849_at (*TCFL5_E8/CHA*) y 235694_at (*TCFL5_E6*) descritas previamente. La expresión de *TCFL5_E8/CHA* fue 7,78 y 4,10 veces mayor en muestras de carcinoma en las colecciones *Skrzypcak Colorectal 2* y *Hong Colorectal*, respectivamente (Figura 11A). La expresión de *TCFL5_E6* también se vio inducida 3,63 veces en la colección *Skrzypcak Colorectal 2* y 1,62 veces en la colección *Hong Colorectal*. Aunque todas las isoformas estudiadas se encontraron inducidas, la mayor variación se produjo en *TCFL5_E8* y/o *CHA* por lo que centramos todo nuestro estudio en estas isoformas. Para profundizar más en la expresión de *TCFL5* en cáncer colorrectal, utilizamos datos de RNA-Seq de adenocarcinoma de colon de la colección *TCGA PanCancer Atlas* [134] utilizando la base de datos *cBioPortal* [135]. De las 267 muestras de esta colección, 104 de ellas (39%) presentaron algún tipo de alteración en *TCFL5* (Figura 11B). Estas alteraciones fueron, en orden de menor a mayor: mutaciones sin sentido, amplificaciones y una sobreexpresión del ARNm. Además, debido a que la expresión de *TCFL5* se ha relacionado con la amplificación cromosómica 20q [116], quisimos estudiar si la expresión de éste se relacionaba con el número de copias del gen (Figura 11C). La correlación de Pearson obtenida fue de 0.69 indicando que la expresión de *TCFL5* depende en gran medida del número de copias que haya de este en el genoma de las células de cáncer de colon.

A continuación, con objeto de decidir el tipo celular sobre el que trabajar, estudiamos la expresión basal de este gen en las líneas de cáncer de colon HT29, HCT116, SW620 y SW480 mediante qPCR (Figura 12A). La línea celular HT29 fue la línea con mayor expresión basal (menor ΔC_t) de *TCFL5/CHA* mientras que las líneas HCT116, SW620 y SW480 presentaron una menor expresión (mayor ΔC_t) sin diferencias significativas entre ellas. Por ello, decidimos utilizar la línea HT29 por ser la que mayor expresión presentó, y la línea HCT116 como representante de las líneas con menor expresión. Mediante amplificación de los transcritos completos presentes en las células HT29 y HCT116 (Figura 12B) observamos que, en la línea HT29 se amplificaron las isoformas *TCFL5_E8/E8b*, *TCFL5_E6/E6b* y *TCFL5_E7*, pero no la isoforma *CHA*. En la línea celular HCT116, se amplificaron las isoformas *TCFL5_E8/E8b* y *CHA* y en menor medida las isoformas *TCFL5_E6/E6b* y *TCFL5_E7*. Además, para descartar la existencia de posibles isoformas que utilicen el exón inicial E2b combinado con los exones finales E6 y E7, se intentaron amplificar

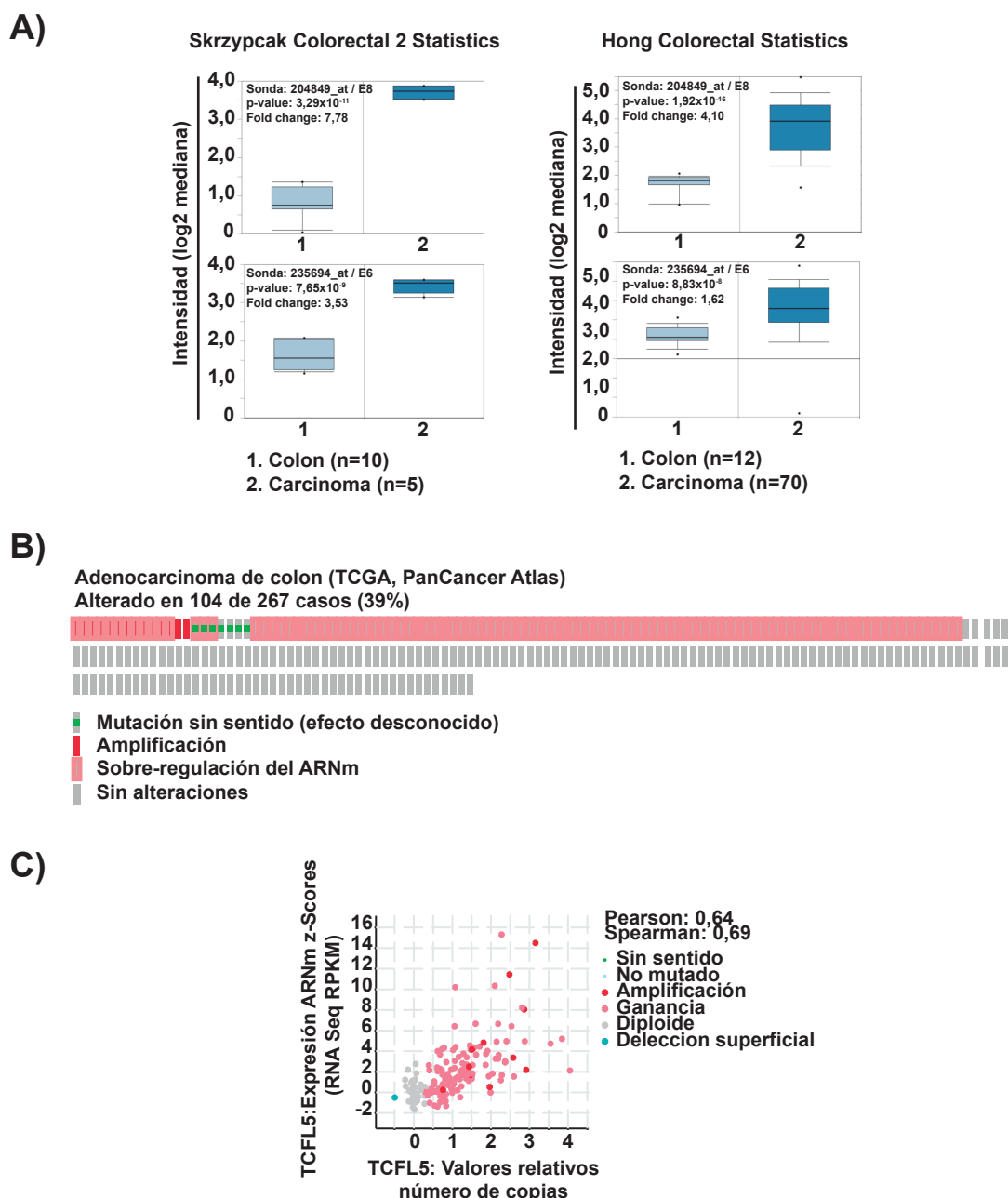


Figura 11. La expresión de *TCFL5* se encuentra elevada en muestras humanas de cáncer de colon. **A)** Análisis de la expresión de *TCFL5* en microarrays de las colecciones independientes de cáncer de colon humano *Skrzypcak colorectal 2* (izquierda) y *Hong colorectal* (derecha) obtenidas de la base de datos *OncoMine* (Thermo Fisher Scientific). Para la detección de la expresión de *TCFL5* se utilizaron dos sondas: 204849_at (*TCFL5_E8/CHA*) y 235694_at (*TCFL5_E6*). **B)** Alteraciones en el ARNm de *TCFL5* por RNA-Seq en la colección “*TCGA, PanCancer Atlas*” de muestras humanas de cáncer de colon obtenidas de la base de datos cBioPortal (Memorial Sloan Kettering Cancer Center). Los casos observados se corresponden con mutaciones sin sentido (verde), amplificaciones del gen (rojo oscuro), sobre-regulación del ARNm (bordeado rojo) y sin alteraciones (gris). **C)** Representación de la expresión del gen frente al número de copias de este. Se representan mutaciones sin sentido (verde), sin mutaciones (azul claro), amplificaciones (rojo oscuro), ganancia (rojo claro), diploidia (gris) y delección superficial (azul oscuro).

estos transcritos. Ni en la línea celular HCT116 ni en HT29 se obtuvo un resultado positivo. Para continuar profundizando en la expresión de las isoformas de *TCFL5* en estas líneas, se amplificaron las uniones de exones descritas previamente (Figura 12C). Ambas líneas celulares presentaron las uniones de exones E1/E2, E2/E3, E3/E4, E4/E5, E5/E7 y E5/E8, sin embargo, únicamente la línea HCT116 presentó las uniones de exones E2b/E3 y E5/E6. Las uniones de exones E6/E7 y E7/E8 no se

amplificaron indicando nuevamente que estas no se producen en los procesos de *splicing* alternativo. En conjunto, estos resultados concuerdan con los anteriores e indican que ambas líneas expresan las isoformas *TCFL5_E8/E8b* (*TCFL5*) y *TCFL5_E7* mientras que la línea celular HT29, a diferencia de las HCT116, carece o tiene una menor expresión de las formas *TCFL5_E6/E6b* y *CHA*. Por último, generamos líneas estables HT29-D6-luc, que expresan luciferasa, y HCT116 para la sobreexpresión de las isoformas *TCFL5* y *CHA* (Figura 12D). La sobreexpresión de las isoformas *TCFL5* y *CHA* produjo una expresión estable en el tiempo en ambas líneas celulares a las que nos referiremos a partir de este momento como Empty Vector (EV), *CHA* y *TCFL5*.

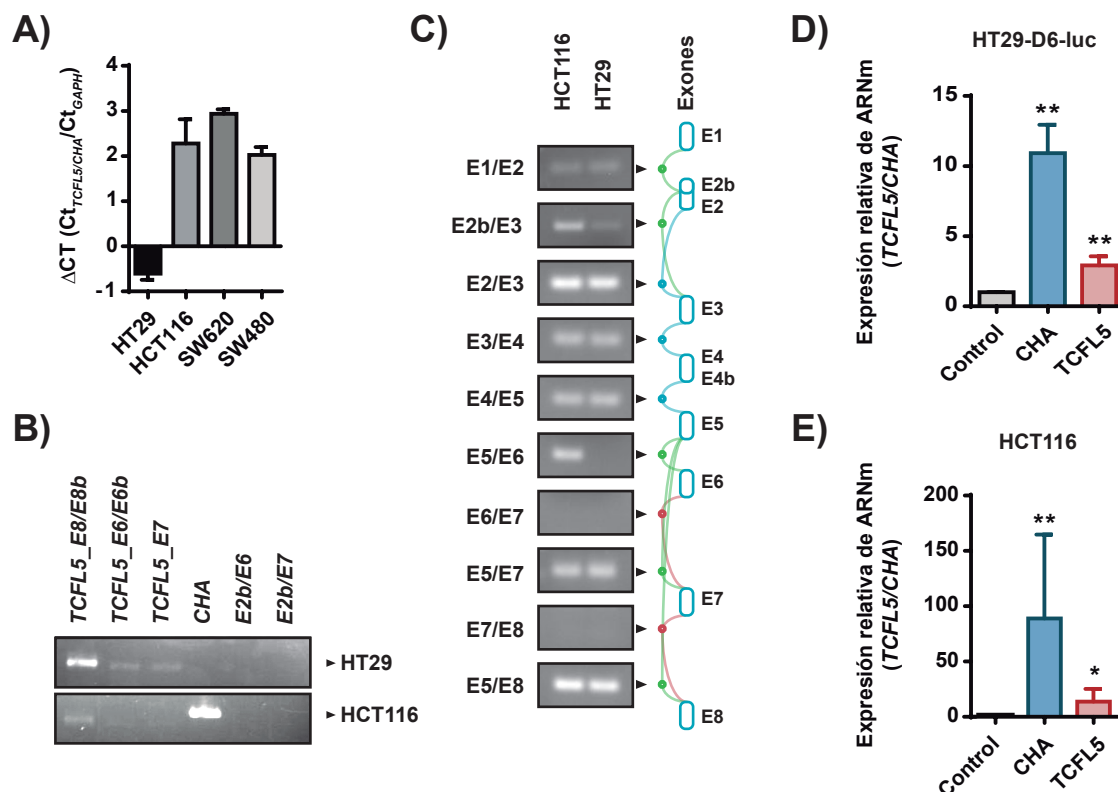


Figura 12. Expresión de las isoformas de *TCFL5* en las líneas de cáncer de colon HT29 y HCT116. **A)** Análisis por qPCR de la expresión de *TCFL5/CHA* en las líneas de cáncer de colon HT29, HCT116, SW620 y SW480. Se representa el ΔCt ($Ct_{TCFL5/CHA} / Ct_{GAPDH}$) de cada línea celular. **B)** Amplificación por PCR de las isoformas completas *TCFL5_E8*, *TCFL5_E8b*, *TCFL5_E6*, *TCFL5_E6b*, *TCFL5_E7* y *CHA* y de las posibles isoformas *E2b/E6* y *E2b/E7* en las líneas de cáncer de colon HT29 y HCT116. **C)** Análisis de la expresión por PCR de las uniones de exones E1/E2, E2b/E3, E2/E3, E4/E5, E5/E6, E6/E7, E5/E7, E7/E8 y E5/E8 en las líneas celulares HT29 y HCT116. **D-E)** Análisis por qPCR de la expresión de *TCFL5/CHA* en las líneas estables HT29-D6-luc (D) y HCT116 (E) control (gris) y con sobreexpresión de *CHA* (azul) y *TCFL5* (rojo). En las gráficas se representa la media \pm SEM de tres experimentos independientes y su estadística respecto al control (* $p < 0,05$; ** $p < 0,01$).

2.2. Efecto de la sobreexpresión de *TCFL5* y *CHA* en el fenotipo tumoral de líneas de cáncer de colon.

2.2.1. La sobreexpresión de *TCFL5* y *CHA* promueve un fenotipo diferente in vitro dependiente de línea celular.

Para estudiar el papel de *TCFL5* y *CHA* en la tumorigénesis de líneas de cáncer de colon, se realizaron diferentes ensayos en las líneas estables HT29-D6-luc y HCT116 generadas para

ver el efecto de la sobreexpresión de TCFL5 y CHA. Como medida de proliferación realizamos un conteo celular durante varios días (Figura 13A) y realizamos ensayos de MTT para medir el metabolismo celular (Figura 13B). También analizamos la capacidad de generar colonias a partir de un bajo número de células (Figura 13C) y estudiamos la migración celular por ensayos de *Wound Healing* en células HCT116, pero no en HT29-D6-luc ya que esta no presentó capacidad migratoria (Figura 13D). En la línea HT29-D6-luc, únicamente la sobreexpresión de CHA produjo efecto aumentando la proliferación, tanto mediante conteo celular como por medición con MTT, y la capacidad clonogénica mientras que TCFL5 no tuvo ningún efecto. Sin embargo, en la línea HCT116, únicamente la sobreexpresión de TCFL5 tuvo efecto disminuyendo la proliferación, tanto por conteo celular como por MTT, y la capacidad clonogénica. No obstante, la migración de las células HCT116 se vio disminuida por la sobreexpresión de CHA, pero no por la de TCFL5.

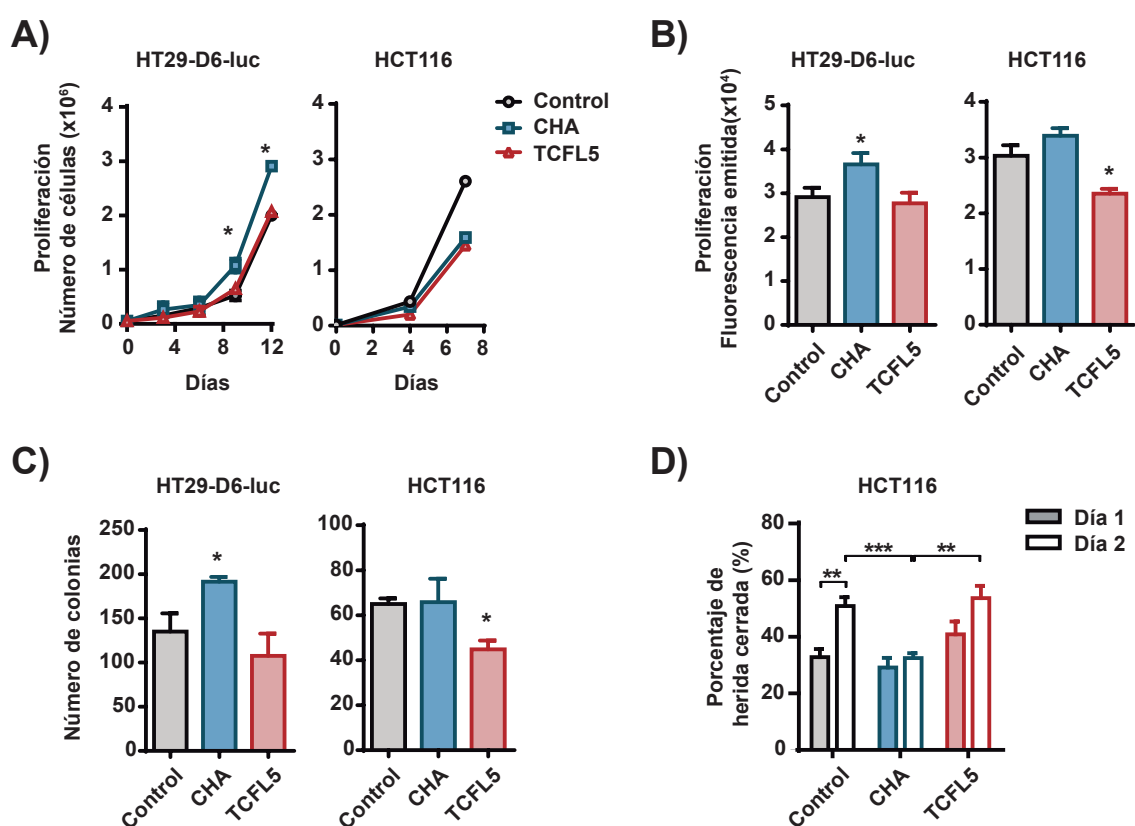


Figura 13. La sobreexpresión de TCFL5 y CHA produce un fenotipo diferente dependiente de la línea celular. **A)** Proliferación por conteo celular a diferentes días de las líneas estables HT29-D6-luc (izquierda) y HCT116 (derecha) control (círculo gris) y que sobreexpresan CHA (cuadrado azul) y TCFL5 (triángulo rojo). **B)** Proliferación celular medida por ensayos de MTT de las líneas estables HT29-D6-luc (izquierda) y HCT116 (derecha) control (gris) y que sobreexpresan CHA (azul) y TCFL5 (rojo). **C)** Capacidad clonogénica medida por número de colonias de las líneas estables HT29-D6-luc (izquierda) y HCT116 (derecha) control (gris) y que sobreexpresan CHA (azul) y TCFL5 (rojo). **D)** Migración celular medida por ensayos de *Wound healing* a diferentes días de la línea estable HCT116 control (gris) y que sobreexpresan CHA (azul) y TCFL5 (rojo). En las gráficas se representa la media \pm SEM de tres experimentos independientes y su estadística (* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$). Los asteriscos sobre la barra indican la estadística respecto al control. Los asteriscos sobre la línea indican la estadística entre las condiciones que une.

2.2.2. La sobreexpresión de TCFL5 afecta al ciclo celular en la fase G2/M.

Con el objetivo de profundizar en el estudio de la proliferación, realizamos ensayos de ciclo celular. Se trataron las líneas celulares estables HCT116 EV, CHA y TCFL5 con colchicina

para sincronizar el ciclo celular en fase G2/M y estudiar por citometría de flujo la entrada a las distintas fases a las 24 horas (Figura 14A). Tras este tiempo, la sobreexpresión de TCFL5 produjo una disminución de células en fase G0/G1 (crecimiento celular) y un aumento de células en fase G2/M (preparación para la mitosis) sin presentar cambios en fase S (síntesis de ADN). Analizamos las fases del ciclo a distintos tiempos durante 48 horas tras el tratamiento para determinar la fase del ciclo que se encontró afectada (Figura 14B). Observamos que las células con sobreexpresión de TCFL5 disminuyeron el número de células en fase G0/G1 a partir de las 12 horas seguido de un aumento de células en fase G2/M y una disminución de células en fase S a las 48 horas. Estos datos sugieren que, mientras que la sobreexpresión de CHA no tiene efecto sobre el ciclo celular en estas

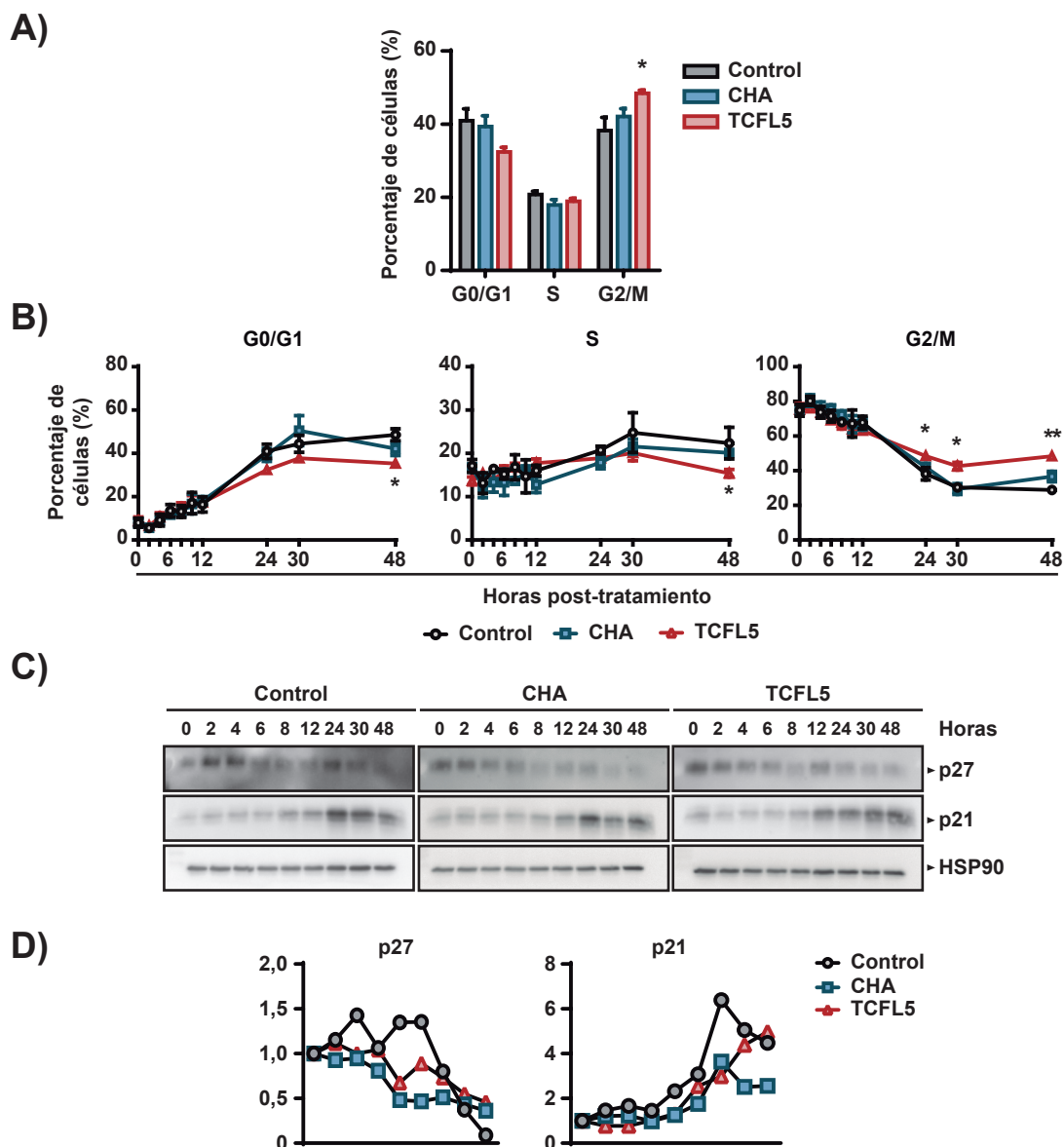


Figura 14. La sobreexpresión de TCFL5 ralentiza el ciclo celular. Estudio del ciclo celular tras el tratamiento con 10nM de colchicina durante 24h en la línea celular estable HCT116 control (gris) y que sobreexpresa CHA (azul) y TCFL5 (rojo). **A)** Porcentaje de células en fase G0/G1, S y G2/M a las 24 h tras el tratamiento. **B)** Porcentaje de células en cada línea en fase G0/G1, S, G2/M y SubG1 a distintos tiempos tras el tratamiento. **C)** Análisis de la expresión por WB de las proteínas implicadas en el ciclo celular p27 y p21 y HSP90 como control de carga. **D)** Cuantificación por densitometría del WB normalizando con la expresión de HSP90. En las gráficas se representa la media \pm SEM de tres experimentos independientes y su estadística (* $p < 0,05$; ** $p < 0,01$)

células, la de TCFL5 favorece la fase G2/M.

Los inhibidores del ciclo celular p27 y p21 son proteínas importantes en el control del ciclo celular y su expresión se asocia a una parada de este [136]. Paralelamente a la citometría, se analizó por WB la expresión de p21 y p27 (Figura 14C) y se cuantificó (Figura 14D). La expresión de p27 disminuyó a lo largo del tiempo siendo esta disminución ligeramente más temprana para las líneas con sobreexpresión de CHA y TCFL5. Por otro lado, la expresión de p21 aumentó en el tiempo en las todas las líneas celulares, aunque con una cinética algo más retardada para las que sobreexpresan TCFL5 mientras que la sobreexpresión de CHA presentó menos inducción de p21 en todo momento. xenoinjertados dependiendo del tipo celular.

2.2.3. La sobreexpresión de TCFL5 y CHA afecta al tamaño de los tumores xenoinjertados dependiendo del tipo celular.

Para terminar de analizar la capacidad tumoral que produce la sobreexpresión de las isoformas TCFL5 y CHA, se inoculaban las líneas estables HT29-D6-luc y HCT116 de manera subcutánea en ratones inmunodeprimidos y se siguió el crecimiento tumoral a lo largo del tiempo.

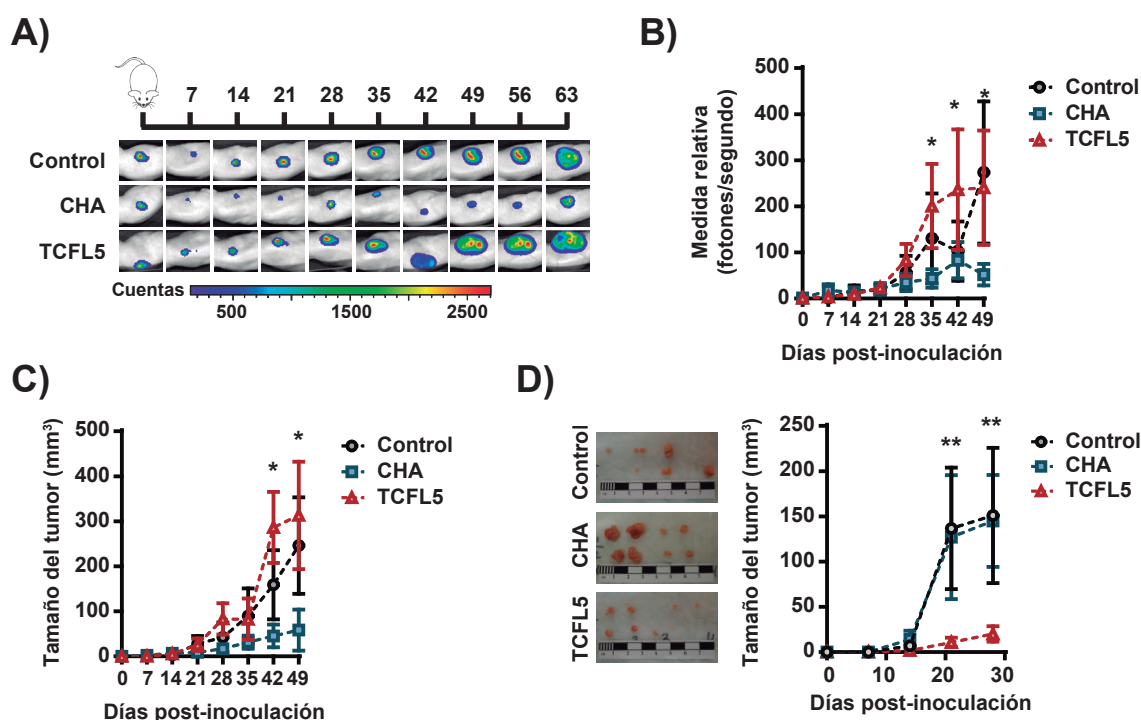


Figura 15. La sobreexpresión de TCFL5 y CHA produce una reducción de los tumores xenoinjertados dependiente de línea celular. Formación de tumores xenoinjertados de las líneas celulares estables HT29-D6-luc y HCT116 control (círculo gris) y que sobreexpresan CHA (cuadrado azul) y TCFL5 (triángulo rojo) tras su inoculación subcutánea en ratones inmunodeprimidos. **A)** Imágenes representativas a lo largo del experimento de la bioluminiscencia emitida por los tumores xenoinjertados de las líneas HT29-D6-luc tras la inoculación subcutánea de luciferina. **B)** Bioluminiscencia media normalizada a día 0 de los tumores formados por las líneas estables HT29-D6-luc a distintos días. **C)** Tamaño medio de los tumores formados por las líneas estables HT29-D6-luc a distintos días medido con calibre milimétrico. **D)** A la izquierda, imagen de los tumores generados por las líneas estables HCT116 a día final del experimento. A la derecha, tamaño medio de los tumores a distintos días medido con calibre milimétrico. En las gráficas se representa la media \pm SEM de dos experimentos independientes para los tumores xenoinjertados de HT29-D6-luc ($n_{\text{ratones total}}=10$ para cada grupo) y un experimento para las HCT116 ($n_{\text{ratones total}}=10$ para cada grupo) y su estadística respecto al control (* $p<0,05$; ** $p<0,01$).

La línea HT29-D6-luc, que expresa luciferasa, nos permitió ver el crecimiento tumoral midiendo la luminiscencia emitida por las células tras la inoculación de luciferina subcutánea (Figura 15A-B). En esta línea celular, que presentó una menor expresión basal de *CHA*, su sobreexpresión produjo un menor crecimiento a lo largo del tiempo de las células con respecto al control. Por su parte, la sobreexpresión de TCFL5 no tuvo ningún efecto significativo comparando con el control. Además, la medida del tamaño del tumor con calibre milimétrico (Figura 15C) confirmó que el menor crecimiento celular observado en las células CHA producía un tumor de menor tamaño al de las células EV. El tamaño del tumor producido por las células TCFL5 fue ligeramente mayor que el control, aunque no significativo. Por último, también medimos con calibre milimétrico los tumores generados por las líneas estables HCT116, que presentaban mayor expresión basal de *CHA* que las HT29-D6-luc (Figura 15D). En este caso, la sobreexpresión de TCFL5 redujo significativamente el tamaño del tumor mientras que la sobreexpresión de CHA no tuvo efecto.

2.3. TCFL5 y CHA regulan su expresión de manera recíproca.

Durante el establecimiento de las líneas celulares observamos que la sobreexpresión de TCFL5 afectaba a la expresión de CHA y viceversa, generando problemas a la hora de interpretar los resultados. Debido a esto nos planteamos la posibilidad de la existencia de una regulación entre isoformas. Analizamos la expresión de *TCFL5* y *CHA* por uniones de exones en las líneas estables HCT116 (Figura 16A). La línea HCT116 EV presentó expresión de las uniones de exones E2b/E3 y E5/E8 indicando expresión de *CHA* endógeno y *TCFL5/CHA* endógeno, respectivamente. La sobreexpresión de TCFL5 produjo una inducción de la unión de exones E5/E8 debido a su propia sobreexpresión exógena, junto con un aumento en la expresión de la unión de exón E2b/E3, indicando una inducción de *CHA* endógeno. Por otro lado, la sobreexpresión de CHA produjo también un aumento de la unión de exones E5/E8 debido a la sobreexpresión de CHA exógeno (que no es detectable mediante la unión de exones E2b/E3 porque amplifica parte del 5'UTR de CHA), mientras que no se observa inducción de la unión de exón E2b/E3, indicando que no hay inducción de *CHA* endógeno. Además, aunque los anticuerpos disponibles en las diferentes casas comerciales no detectan de manera eficiente la expresión basal de TCFL5 por Western-blot en las líneas celulares utilizadas durante este estudio, quisimos analizar la expresión de estas isoformas a nivel de proteína ya que si se puede observar expresión cuando hay una alta sobreexpresión de estas (Figura 16B). De esta forma observamos que en la línea HCT116 CHA hay una sobreexpresión de TCFL5. Sin embargo, en la línea HCT116 TCFL5 no se observan variaciones en CHA. Estos datos juntos sugieren que podría existir una regulación entre isoformas de tal forma que la sobreexpresión de TCFL5 conlleva una sobreexpresión de CHA y viceversa.

Finalmente, para estudiar si la posible regulación TCFL5/CHA se debe a la unión directa de estos a su propio promotor, llevamos a cabo una inmunoprecipitación de cromatina para cada una de las isoformas. Transfectamos células HCT116 con las isoformas TCFL5 y CHA fusionadas

a la etiqueta de reconocimiento FLAG para una inmunoprecipitación específica. Se analizaron por qPCR dos regiones adyacentes al inicio de transcripción de *TCFL5* (TCFL5_01 y TCFL5_03) y dos regiones adyacentes al inicio de transcripción de *CHA* (CHA_01 y CHA_02). Como control negativo se amplificaron zonas cercanas al promotor pero transcripcionalmente inactivas (control_01, control_02 y control_03) (Figura 16C). Tras el análisis, observamos una ligera amplificación de las regiones control lo que indica un ligero fondo en la inmunoprecipitación. Con respecto al promotor de *TCFL5*, tanto TCFL5 como CHA se unieron a las regiones promotoras TCFL5_01 y TCFL5_02. En cuanto al promotor de *CHA*, tanto TCFL5 como CHA se unieron a la región CHA_02 pero no a la región CHA_01. Estos resultados indican que estos factores de transcripción se unen a su propio promotor y por tanto podrían regular su propia expresión directamente.

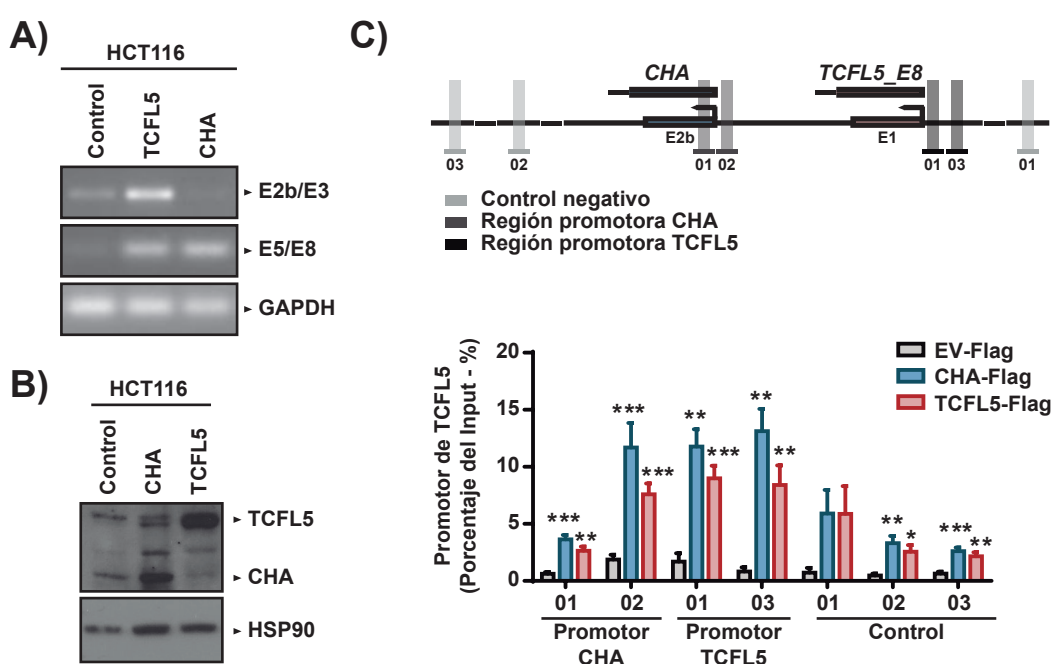


Figura 16. TCFL5 y CHA regulan su expresión de manera recíproca. **A)** Expresión por PCR de las uniones de exones E2b/E3 endógena y E5/E8 y GAPDH en las líneas celulares estables HCT116 control y que sobreexpresan TCFL5 y CHA. **B)** Análisis por Western-blot de la expresión de proteína de TCFL5, CHA y HSP90 como control de carga, en las líneas celulares estables HCT116 control y que sobreexpresan TCFL5 y CHA. **C)** Arriba, esquema de las regiones control y del promotor de *TCFL5* y *CHA* estudiadas en los experimentos de CHIP. Abajo, análisis por qPCR de los ensayos de CHIP de la unión de TCFL5 y CHA a las regiones control y a su promotor en la línea celular HCT116 transfectada con los plásmidos Control-Flag (gris), CHA-Flag (azul) y TCFL5-Flag (rojo). En las gráficas se representa la media \pm SEM de tres experimentos independientes y su estadística respecto a la transfección EV-Flag de cada región estudiada (* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$).

2.4. Descripción del efecto de la deficiencia de TCFL5 y CHA en el fenotipo tumoral de líneas de cáncer de colon.

Los resultados obtenidos de la caracterización del fenotipo de células de cáncer de colon con sobreexpresión de TCFL5 y CHA indican que estos factores de transcripción presentan un papel en el proceso tumoral, diferente el uno del otro, y dependiente de línea celular. Por un lado, la sobreexpresión de CHA favorece la proliferación celular, la capacidad clonogénica y disminuye la

formación de tumores en células HT29, mientras que la de TCFL5 no tiene ningún efecto. Por otra parte, en líneas HCT116 la sobreexpresión de TCFL5 disminuye la proliferación celular, la capacidad clonogénica y la formación de tumores mientras que CHA disminuye la migración. Además, estos factores de transcripción se unen a su propio promotor pudiendo controlar así su propia expresión. Dada la complejidad de interpretación de los resultados debido a la posible regulación recíproca de estos factores de transcripción, se continuó el estudio con el silenciamiento de todas las isoformas de TCFL5 presentes en ambas líneas celulares.

2.4.1. El silenciamiento de TCFL5/CHA produce cambios en el fenotipo de las líneas de cáncer de colon.

Se generaron líneas estables HT29 y HCT116 para el silenciamiento de TCFL5 y CHA mediante shARN. Para ello se utilizaron los vectores lentivirales scramble (SCR) como control, shTCFL5_12 (que mapea en el exón E3 de TCFL5), shTCFL5_29 (que mapea en el exón E8) y shTCFL5_90 (que mapea en el exón E2 inmediatamente delante de la metionina iniciadora de la región codificante de CHA) (Figura 17A). En la línea HT29 el silenciamiento produjo una reducción de la expresión de *TCFL5/CHA* significativa mientras que en la línea HCT116 dicho silenciamiento no fue eficiente por lo que no se utilizaron (datos no mostrados).

En la línea HT29, los silenciamientos tuvieron distintos efectos en proliferación (Figura 17C). La línea shTCFL5_90 presentó un crecimiento mayor que la línea SCR mientras que en la

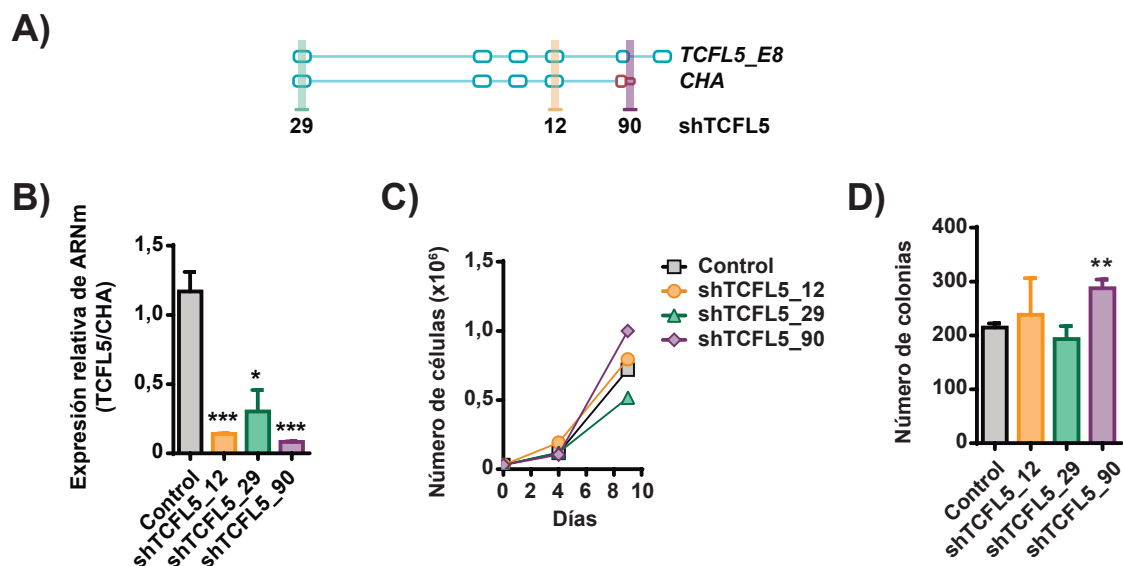


Figura 17. El silenciamiento de TCFL5/CHA produce un fenotipo dependiente de línea celular. **A)** Esquema de las regiones de reconocimiento en la secuencia de las isoformas TCFL5 y CHA de los silenciamientos shTCFL5_12 (amarillo), shTCFL5_29 (verde) y shTCFL5_90 (morado). **B)** Análisis de la expresión por qPCR de *TCFL5/CHA* en las líneas generadas HT29 estables control (gris) y con los silenciamientos shTCFL5_12 (amarillo), shTCFL5_29 (verde) y shTCFL5_90 (morado). **C)** Proliferación por conteo celular a diferentes días de las líneas estables HT29 control (cuadrado negro), shTCFL5_12 (círculo amarillo), shTCFL5_29 (triángulo verde) y shTCFL5_90 (rombo morado). **D)** Capacidad clonogénica medida por número de colonias de las líneas estables HT29 control (gris), shTCFL5_12 (amarillo), shTCFL5_29 (verde) y shTCFL5_90 (morado). En las gráficas se representa la media \pm SEM de tres experimentos independientes y su estadística respecto al control (* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$).

línea shTCFL5_29 fue menor. La línea shTCFL5_12 no presentó diferencias respecto a la línea SCR. La capacidad clonogénica (Figura 17D) solo la línea shTCFL5_90 presentó un mayor número de colonias.

2.4.2. La deficiencia de TCFL5/CHA promueve un fenotipo menos agresivo.

Debido a que los shARNs utilizados mapean en distintas regiones de la secuencia de TCFL5 pudiendo afectar de manera diferente al silenciamiento de las isoformas del gen y junto a la rápida pérdida de dicho silenciamiento estable, los resultados obtenidos fueron difícilmente interpretables y poco reproducibles. Por ello se generaron, gracias a la técnica de edición génica CRISPR-Cas9, líneas HCT116 deficientes en la expresión de TCFL5 y CHA (HCT116 TCFL5^{-/-}) tras la delección del exón 3 alterando la pauta de lectura y generando señales de parada de la traducción produciendo proteínas no funcionales (Figura 18A). Se generaron dos clones HCT116 TCFL5^{-/-} deficientes en la expresión de TCFL5 y CHA a los se referirá como KO23 y KO34 y un clon control HCT116 TCFL5^{+/+} con una expresión normal de TCFL5 y CHA que se nombrará como WT6. Se amplificó la región del ADN comprendida entre los exones E2b y E4 para comprobar la delección del exón 3 (Figura 18B). Se observó que el clon WT6 obtuvo un amplicón de igual tamaño que la línea celular HCT116 correspondiente al gen normal, mientras que los clones KO23 y KO34 presentaron un amplicón de menor tamaño correspondiente con la delección del exón 3. Dicho amplicón fue secuenciado para

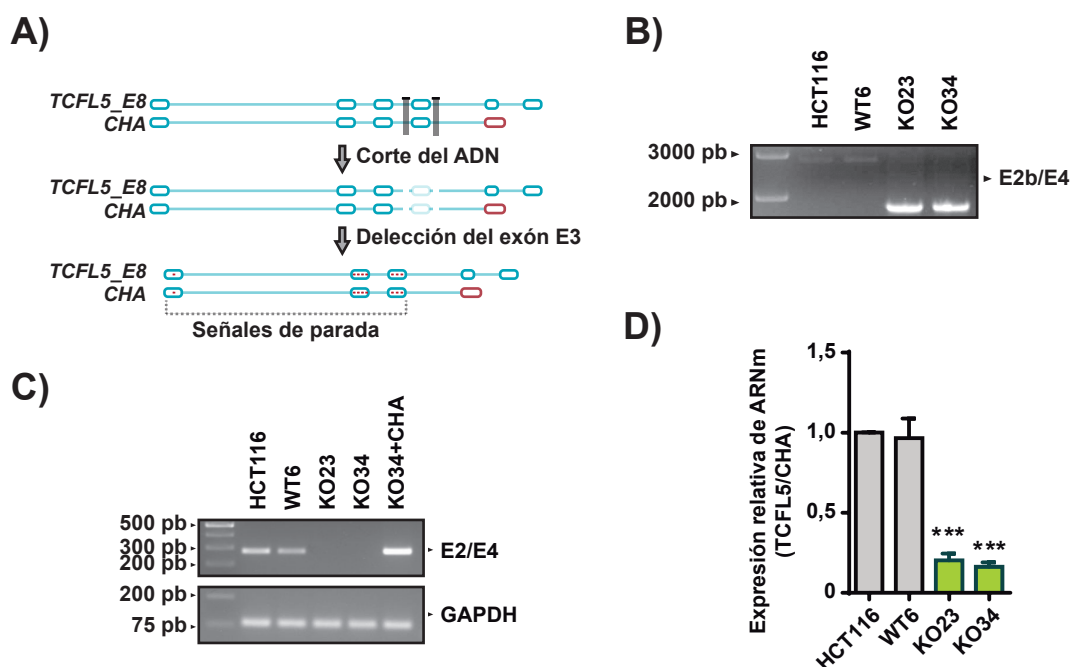


Figura 18. Generación de líneas HCT116 TCFL5^{-/-} mediante *Self-cloning* CRISPR-Cas9. **A)** Esquema de la estrategia de delección del exón 3 de *TCFL5* utilizando dos guías para la generación de células HCT116 TCFL5^{-/-}. **B)** Amplificación por PCR de la región E2b/E4 del ADN de células HCT116, clones TCFL5^{+/+} (WT6) y clones TCFL5^{-/-} (KO23 y KO34). **C)** Amplificación por PCR de la región E2/E4 del ARNm de células HCT116, clones TCFL5^{+/+} (WT6) y clones TCFL5^{-/-} (KO23 y KO34). **D)** Análisis de la expresión por qPCR de *TCFL5/CHA* de células HCT116 (gris), clones TCFL5^{+/+} (WT6; gris) y clones TCFL5^{-/-} (KO23 y KO34; verde). En las gráficas se representa la media ± SEM de tres experimentos independientes y su estadística respecto al control (***)p<0,01).

comprobar la correcta delección (datos no mostrados). Además, se amplificó la región comprendida entre los exones E2 y E4 en el ARNm de estas líneas para determinar la expresión del exón 3 (Figura 18C). El clon WT6 expresó el exón 3 de la misma forma que la línea HCT116 mientras que los clones KO23 y KO34 no lo expresaron. La transfección de CHA en la línea KO34 recuperó la expresión del exón E3. Por último, analizamos la expresión de la unión de exones E5/E8 por qPCR (Figura 18D). El clon WT6 presentó una expresión de *TCFL5* y *CHA* igual que la línea HCT116. Sin embargo, los clones KO23 y KO34, presentaron únicamente una disminución en la expresión de esta unión de exones debido a una transcripción residual de los últimos exones.

Tras la generación de las líneas HCT116 *TCFL5*^{-/-} se continuó con el estudio del papel que produce la deficiencia de *TCFL5*/*CHA* en el proceso tumoral de cáncer de colon. Las líneas KO23 y KO34, deficientes en la expresión de *TCFL5*/*CHA*, presentaron una menor proliferación con respecto a la línea WT6 (Figura 19A) así como una menor actividad metabólica medida por ensayos de MTT

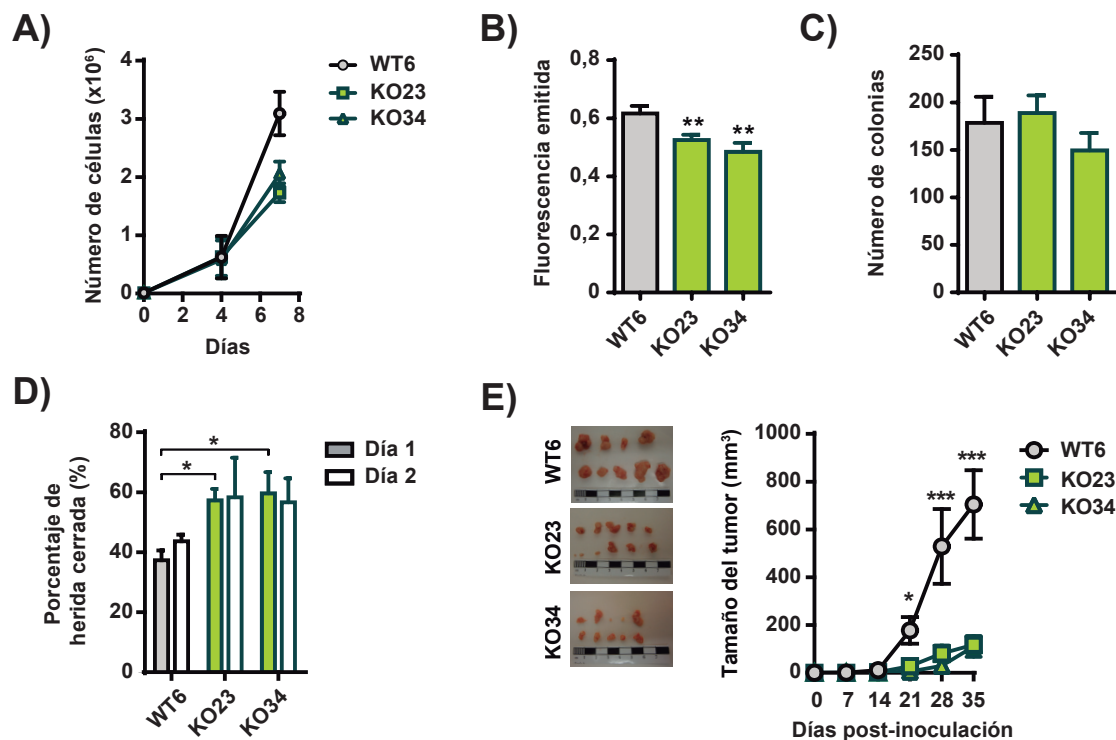


Figura 19. La deficiencia de *TCFL5* y *CHA* promueve un fenotipo menos tumoral en células HCT116. **A)** Proliferación por conteo celular de la línea HCT116 *TCFL5*^{+/+} clon WT6 (círculo gris) y las líneas HCT116 *TCFL5*^{-/-} clon KO23 (cuadrado verde) y clon KO34 (triángulo verde). **B)** Proliferación celular medida por ensayos de MTT en la línea HCT116 *TCFL5*^{+/+} clon WT6 (gris) y las líneas HCT116 *TCFL5*^{-/-} clon KO23 (verde) y clon KO34 (verde). **C)** Capacidad clonogénica medida por número de colonias en la línea HCT116 *TCFL5*^{+/+} clon WT6 (gris) y las líneas HCT116 *TCFL5*^{-/-} clon KO23 (verde) y clon KO34 (verde). **D)** Migración celular medida por ensayos de *Wound healing* a diferentes días en la línea HCT116 *TCFL5*^{+/+} clon WT6 (gris) y las líneas HCT116 *TCFL5*^{-/-} clon KO23 (verde) y clon KO34 (verde). **E)** Formación de tumores xenoinjertados de la línea HCT116 *TCFL5*^{+/+} clon WT6 (círculo gris) y las líneas HCT116 *TCFL5*^{-/-} clon KO23 (cuadrado verde) y clon KO34 (triángulo verde) tras su inoculación subcutánea en ratones inmunodeprimidos. A la izquierda, imagen de los tumores generados a día final del experimento. A la derecha, tamaño medio de los tumores a distintos días medido con calibre milimétrico. En las gráficas se representa la media \pm SEM de tres experimentos independientes y un experimento para el caso de los tumores xenoinjertados ($n_{\text{ratones total}}=10$ para cada grupo). Se muestra la estadística (* $p<0,05$; ** $p<0,01$; *** $p<0,001$). Los asteriscos sobre la barra indican la estadística respecto al control. Los asteriscos sobre la línea indican la estadística entras las condiciones que une.

(Figura 19B). Con respecto a la capacidad clonogénica, la deficiencia de TCFL5/CHA de las líneas KO23 y KO34 no conllevó diferencias significativas frente a la línea WT6 (Figura 19C). Las líneas KO23 y KO34 produjeron una mayor migración en ensayos de *Wound healing* con respecto a la línea WT6 (Figura 19D). Por último, se generaron tumores xenoinjertados de las líneas WT6, KO23 y KO34 en ratones inmunodeficientes y se siguió el tamaño tumoral en el tiempo (Figura 19E). Las líneas KO23 y KO34 produjeron tumores de tamaño significativamente más pequeño que la línea WT6.

2.4.3. La deficiencia de TCFL5/CHA aumentó el porcentaje de células en fase G0/G1.

Se analizó el ciclo celular tras la sincronización en fase G2/M en las líneas HCT116 TCFL5^{-/-}. A las 24 horas de tratamiento, el número de células en todas las fases del ciclo no se vio alterada por la deficiencia de TCFL5/CHA (Figura 20A). Sin embargo, la cinética reveló un aumento del número de células en fase G0/G1 a las 48 horas del inicio del tratamiento y una disminución de células en fase S sin alterarse el número de células en fase G2/M (Figura 20B).

Paralelamente se analizó la expresión de los inhibidores de ciclo celular p27 y p21 por WB (Figura 20C) y se cuantificó (Figura 20D). La expresión de p27 disminuyó en el tiempo mientras que la expresión de p21 aumentó. La deficiencia de las isoformas no generó cambios significativos. No obstante, aunque son datos preliminares, sugieren que la deficiencia de las isoformas TCFL5 y CHA produce la parada del ciclo celular en fase G0/G1.

2.5. Análisis de los genes diana y proteínas de interacción de las isoformas TCFL5 y CHA.

Después de estudiar el fenotipo producido por las isoformas TCFL5 y CHA, quisimos analizar a nivel molecular las razones de estas diferencias.

2.5.1. TCFL5 y CHA inhiben la transcripción de *BCL2*.

Para definir dianas potenciales de TCFL5 y/o CHA, realizamos una búsqueda *in silico* de experimentos de CHIP-Seq para TCFL5. Se obtuvieron datos de un experimento de CHIP-Seq realizado en células Lovo de cáncer de colon [137]. Su análisis reveló que *SYK*, *BCL2*, *DUX4*, *HDAC4*, *PTX1* y *DDX51* podían ser potenciales candidatos a genes diana de TCFL5 y/o CHA.

BCL2 es una proteína fundamental en el proceso de apoptosis con un papel antiapoptótico. Su expresión suele ser elevada en distintos tipos de cáncer y generalmente se asocia con un aumento en la resistencia a fármacos y radioterapia [138]. Debido a esto, nos centramos en este gen como diana de transcripción de TCFL5 y/o CHA. Para comprobar los datos obtenidos *in silico* y verificar que *BCL2* es diana transcripcional de TCFL5 y/o CHA, realizamos ensayos de inmunoprecipitación de cromatina. Transfectamos células HCT116 con las distintas isoformas fusionadas a la etiqueta

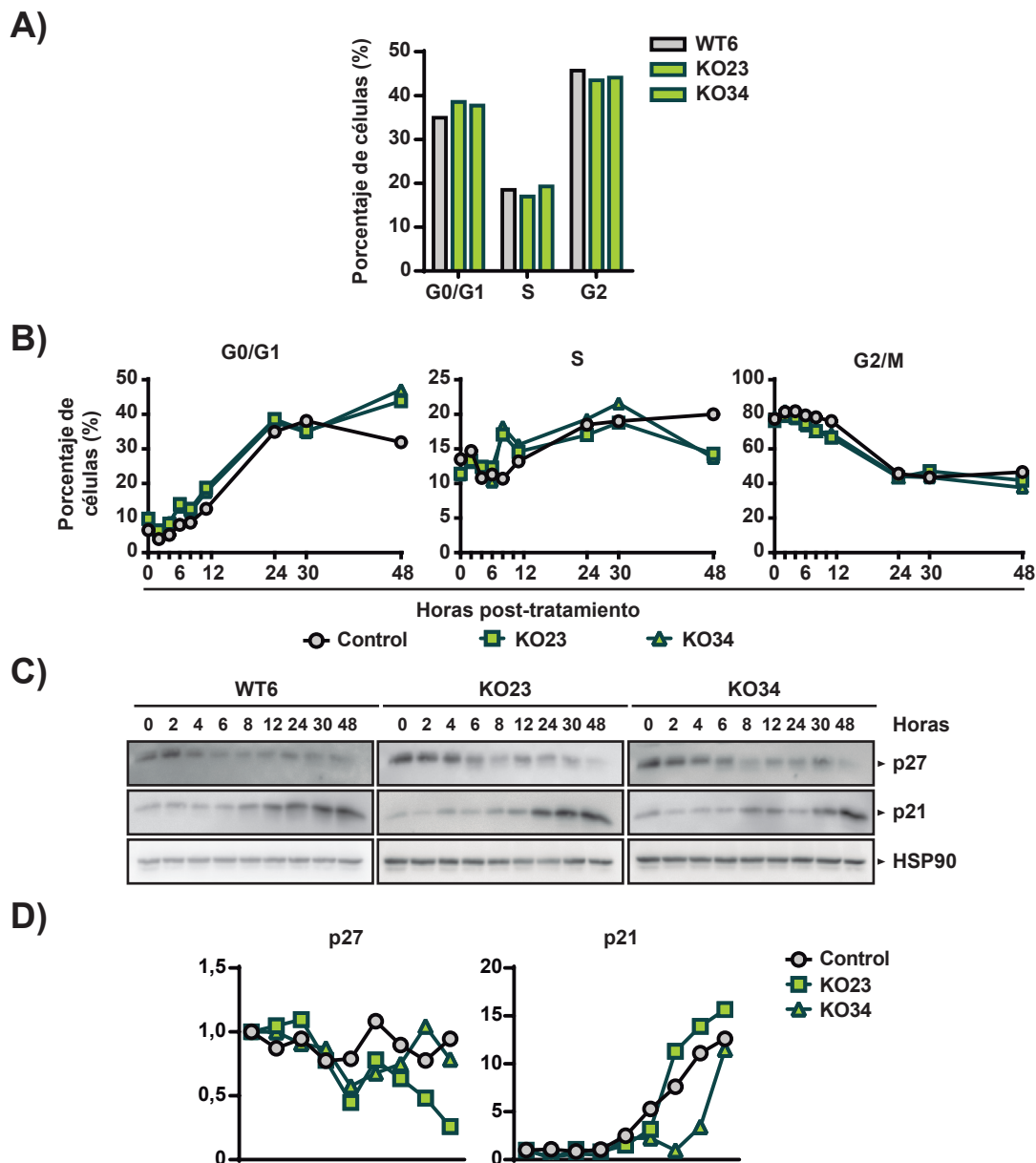


Figura 20. La deficiencia de TCFL5/CHA ralentiza el ciclo celular. Estudio del ciclo celular tras el tratamiento con 10nM de colchicina durante 24h en la línea celular HCT116 TCFL5^{+/+} clon WT6 (círculo gris) y las líneas HCT116 TCFL5^{-/-} clon KO23 (cuadrado verde) y clon KO34 (triángulo verde). **A)** Porcentaje de células en fase G0/G1, S y G2/M a las 24 h tras el tratamiento. **B)** Porcentaje de células en cada línea en fase G0/G1, S, G2/M y SubG1 a distintos tiempos tras el tratamiento. **C)** Análisis de la expresión por Western-Blot de las proteínas implicadas en el ciclo celular p27 y p21 y HSP90 como control de carga. **D)** Cuantificación por densitometría del WB normalizando con la expresión de HSP90.

FLAG para una inmunoprecipitación específica. Se amplificaron por qPCR dos regiones en el promotor de *BCL2*: una región adyacente al inicio de transcripción (positivo) y otra en una región transcripcionalmente inactiva (negativo) (Figura 21A). Tanto TCFL5 como CHA se unieron específicamente a la región promotora de *BCL2*, aunque TCFL5 lo hizo de manera más eficiente. No hay unión de ambas isoformas a la región transcripcionalmente inactiva (Figura 21B). Además, comprobamos la expresión de *BCL2* en las líneas que sobreexpresan TCFL5 y CHA y en las líneas HCT116 TCFL5^{-/-} (Figura 21B). La sobreexpresión de TCFL5 y de CHA redujo prácticamente a la mitad la expresión de *BCL2* mientras que la deficiencia de TCFL5 no tuvo efecto. Estos datos corroboran que *BCL2* es una diana transcripcional de TCFL5 y de CHA y que estos son represores

de su expresión en células HCT116.

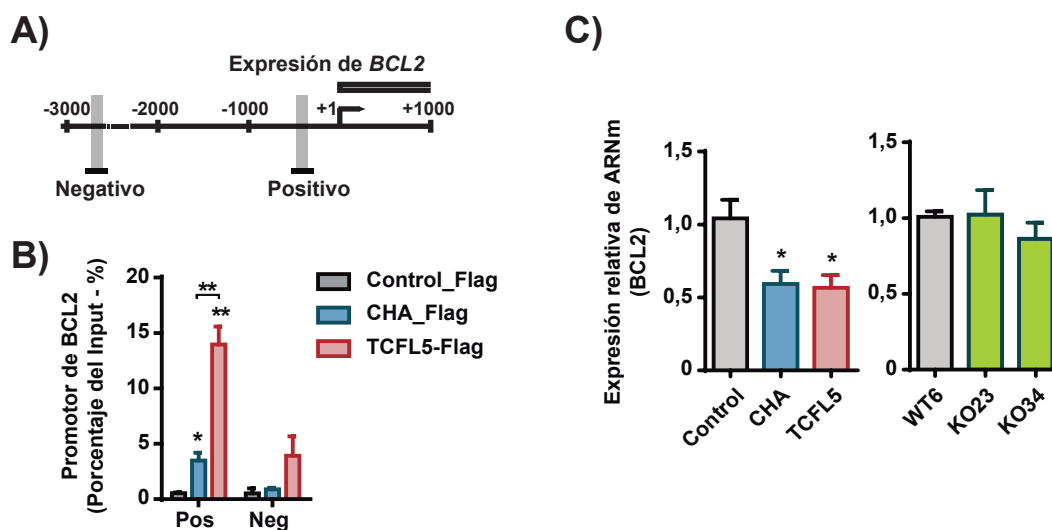


Figura 21. *BCL2* es una diana transcripcional de TCFL5 y CHA, y regulan su expresión. **A)** Esquema del promotor de *BCL2* y las regiones positiva y negativa estudiadas en los experimentos de CHIP. **B)** Análisis por qPCR de los ensayos de CHIP de la unión de TCFL5 y CHA al promotor de *BCL2* en la línea celular HCT116 transfectadas con los plásmidos Control-Flag (gris), CHA-Flag (azul) y TCFL5-Flag (rojo). **C)** Análisis de la expresión por qPCR de *BCL2*. A la izquierda, expresión en líneas celulares estables HCT116 control (gris) y que sobreexpresan CHA (azul) y TCFL5 (rojo). A la derecha, en la línea HCT116 TCFL5^{+/+} clon WT6 (gris) y las líneas HCT116 TCFL5^{-/-} clon KO23 (verde) y clon KO34 (verde). En las gráficas se representa la media \pm SEM de tres experimentos independientes y su estadística (* $p < 0,05$; ** $p < 0,01$). Los asteriscos sobre la barra indican la estadística respecto al control. Los asteriscos sobre la línea indican la estadística entre las condiciones que une.

2.5.2. TCFL5 interacciona con la subunidad p100 de NFKB2 y regula su actividad.

Con el objetivo de identificar proteínas de interacción con las cuales las distintas isoformas de TCFL5 podrían estar formando dímeros o complejos proteicos, realizamos un interactoma de cada una de ellas. Transfectamos células HCT116 con las isoformas fusionadas a la etiqueta FLAG para realizar una inmunoprecipitación específica de cada isoforma. El inmunoprecipitado fue analizado por espectrofotometría de masas (Tabla S3) y los datos filtrados por número de péptidos y *score* con respecto al control (Tabla S4). Los resultados obtenidos mostraron que la red de interacción de la isoforma CHA es más parecida a la de la isoforma TCFL5_E8 que a la del resto, previsiblemente debido a la presencia del exón E8 (Figura S3). Las isoformas TCFL5_E8, TCFL5_E7 y TCFL5_E6, que comparten el exón E1, fueron más parecidas entre sí. No obstante, las redes de interacción de cada una de las isoformas fueron notablemente diferentes. Para TCFL5 encontramos como posibles proteínas de interacción entre otras a la subunidad p100 de NFKB2, DDB1 y PCNA mientras que para el caso de CHA, encontramos como posible proteína de interacción al también factor de transcripción YBX1.

NFKB2 es un factor de transcripción ubicuo que está involucrado en procesos de inflamación, respuesta inmune y supervivencia celular. Las dianas transcripcionales de NFKB2 son diversas y entre ellas se encuentra *BCL2*, uniéndose a su promotor y regulando su expresión [139]. Debido a que *BCL2* es también una diana transcripcional de TCFL5 y CHA, quisimos profundizar en la

posible interacción de TCFL5 con NFKB2. Para certificar que TCFL5 interacciona con la subunidad p100 de NFKB2 realizamos una inmunoprecipitación de p100 en células HCT116 transfectadas con las isoformas TCFL5 y CHA fusionadas a la etiqueta de reconocimiento FLAG. (Figura 22A). Los resultados obtenidos mostraron la co-inmunoprecipitación únicamente de p100 con TCFL5, pero no con CHA, indicando que sólo TCFL5 presenta interacción con la subunidad p100 de NFKB2.

Una vez comprobado que TCFL5 se une a la subunidad p100 de NFKB2, estudiamos si TCFL5 podría estar regulando la actividad NFKB2. Para ello, activamos la vía de NFKB2 utilizando la citoquina TWEAK, la cual pertenece a la familia de ligandos del factor de necrosis tumoral (TNF). El tratamiento con TWEAK conlleva una activación específica de la vía de NFKB2 [140] produciendo, en líneas de cáncer de colon, una inducción débil de apoptosis vía NFKB2 [141] [142]. Tratamos las células HCT116 TCFL5^{-/-} con TWEAK y analizamos por Western-blot la presencia de p100 y p52 (Figura 22B). El clon WT6 y la línea HCT116 presentaron una inducción de p52 en respuesta a TWEAK indicando una similar activación de NFKB2. Las líneas KO23 y KO34 presentaron niveles mucho más bajos de p100 y p52 tras la activación por TWEAK que la WT. Sin

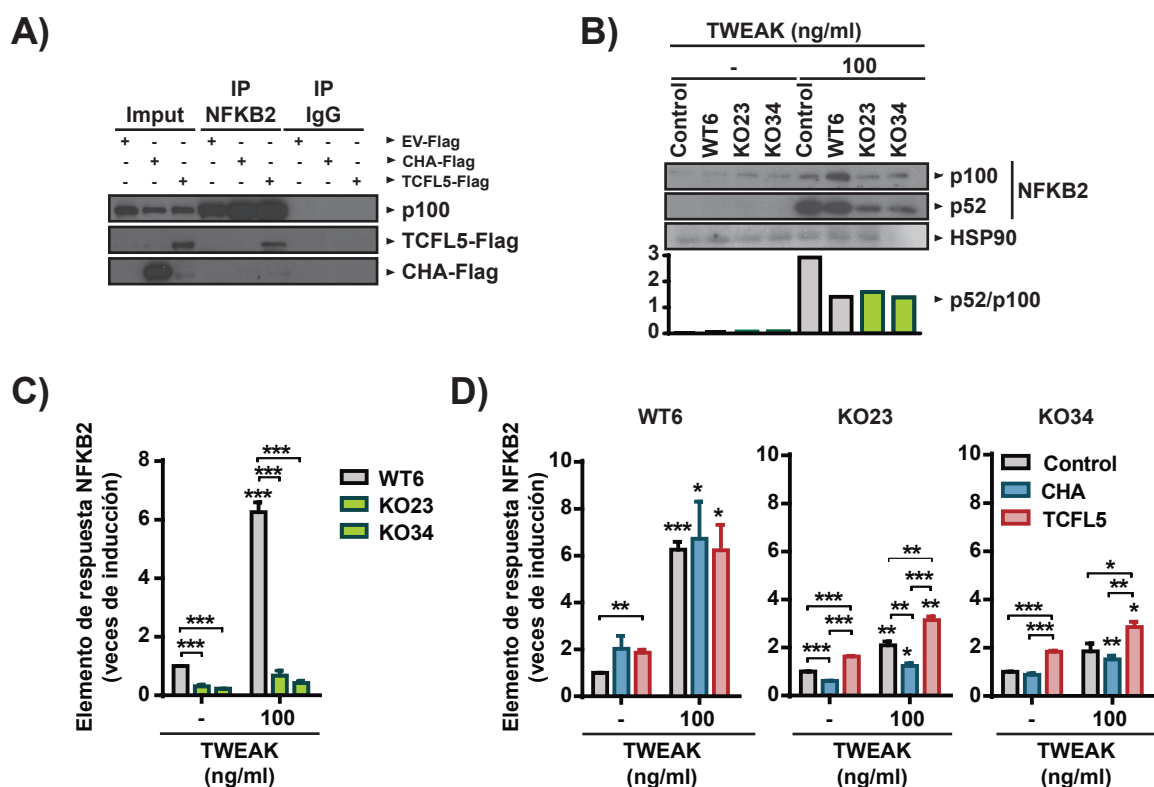


Figura 22. TCFL5 interacciona con NFKB2 y reprime su actividad. **A)** Análisis de la interacción de p100 con TCFL5 mediante inmunoprecipitación. **B)** Arriba, análisis por Western-blot de la activación de NFKB2 tras el tratamiento con 100 ng/ml de TWEAK durante 24 horas en células HCT116 (gris), HCT116 TCFL5^{+/+} clon WT6 (gris) y HCT116 TCFL5^{-/-} clon KO23 (verde) y clon KO34 (verde). Abajo, cuantificación por densitometría del WB normalizando con la expresión de HSP90. **C)** Análisis de la actividad de NFKB2 por actividad luciferasa de su reportero tras el tratamiento 24h con 100 ng/ml de TWEAK de células HCT116 TCFL5^{+/+} clon WT6 (gris) y HCT116 TCFL5^{-/-} clon KO23 (verde) y clon KO34 (verde). **D)** Análisis de la actividad de NFKB2 por actividad luciferasa de su reportero tras el tratamiento 24h con 100 ng/ml de TWEAK en células HCT116 TCFL5^{+/+} clon WT6 y HCT116 TCFL5^{-/-} clon KO23 y clon KO34 transfectadas con los plásmidos Control-Flag (gris), CHA-Flag (azul) y TCFL5-Flag (rojo). En las gráficas se representa la media \pm SEM de tres experimentos independientes y su estadística (* p <0,05; ** p <0,01; *** p <0,001). Los asteriscos sobre la barra indican la estadística respecto al control. Los asteriscos sobre la línea indican la estadística entras las condiciones que une.

embargo, el ratio p100/p52 tras la activación con TWEAK fue igual para las 3 líneas sugiriendo que la deficiencia de TCFL5 y CHA no altera per se la degradación de NFkB2 sino que disminuye los niveles totales de este.

A continuación, utilizamos un ensayo de actividad de NFkB2 mediante un reportero de actividad luciferasa bajo el control de 3 zonas de reconocimiento de NFkB2 en tándem. Transfectamos dicho promotor en células HCT116 y cuantificamos la actividad luciferasa a las 24 horas tras el tratamiento con TWEAK (Figura 22C). En condiciones basales, las líneas KO23 y KO34 presentaron una menor actividad NFkB2 que la línea control WT6. Tras el tratamiento con TWEAK, se produjo una fuerte inducción de la actividad NFkB2 en WT6 siendo significativamente mucho menor en las líneas KO23 y KO34. Además, quisimos estudiar el efecto individual de las isoformas TCFL5 y CHA. Para ello las líneas HCT116 TCFL5^{-/-} se transfectaron con las isoformas TCFL5 y CHA con el objetivo de recuperar la expresión específica de cada una de ellas y realizamos los mismos experimentos de actividad luciferasa (Figura 22D). En la línea WT6, la actividad basal de NFkB2 fue mayor tras la sobreexpresión de TCFL5 y CHA, aunque únicamente significativa para el caso de TCFL5. Por el contrario, el tratamiento con TWEAK indujo de igual forma la actividad NFkB2 en las tres condiciones. En el caso de las líneas KO23 y KO34, en condiciones basales, la transfección de CHA redujo la actividad NFkB2 mientras que TCFL5 la indujo en ambas líneas. Tras el tratamiento con TWEAK, la activación de NFkB2 continuó siendo menor con la sobreexpresión de CHA mientras que aumentó con la sobreexpresión de TCFL5 respecto al control. Los resultados obtenidos indican que TCFL5 favorece la actividad NFkB2 mientras que CHA tiene el efecto contrario.

3. Papel de las isoformas TCFL5 y CHA en la regulación de genes de pluripotencia.

TCFL5 se induce en la formación de esferoides tumorales multicelulares (MCTS) de varias líneas celulares de glioblastoma y en la línea de cáncer de colon HT29. Además, el silenciamiento mediante siARN de *TCFL5* inhibe la formación de estos MCTS reduciendo su número [111]. La formación de MCTS produce la diferenciación de ciertas células a células denominadas *Cancer Stem Cells* (CSCs). Estas células se caracterizan por tener una capacidad de pluripotencia más elevada y aumentar su agresividad teniendo por tanto una mayor capacidad tumoral. La detección de estas células se ha puesto de manifiesto como un tema relevante para la lucha contra el cáncer [143]. Debido a esto, quisimos estudiar si *TCFL5* juega un papel en la formación de CSCs.

3.1. *TCFL5* y *CHA* se encuentran inducidos en los MCTS y controlan su formación.

En primer lugar, quisimos verificar que durante la formación de MCTS de líneas celulares de cáncer de colon hay una inducción en la expresión de *TCFL5*. Generamos MCTS de las líneas celulares HT29, HCT116, SW480 y SW620 y analizamos por qPCR la expresión de *TCFL5/CHA* (Figura 23A). Las células HT29 y HCT116 presentaron una inducción significativa en la expresión

de estas isoformas mientras que en la línea SW620 la inducción no fue significativa y en la línea SW480 no se vio ninguna inducción. Estos datos corroboran la inducción de *TCFL5/CHA* en la formación de MCTS de líneas de cáncer de colon. Para concretar las isoformas inducidas en este proceso, analizamos por qPCR las uniones de exones E3/E4, E5/E6, E5/E7 y E5/E8 en los MCTS de la línea celular HCT116 a los días 7 y 10 de formación de MCTS (Figura 23B). Todas las uniones de exones, a excepción de la unión E5/E6, presentaron una inducción significativa indicando una

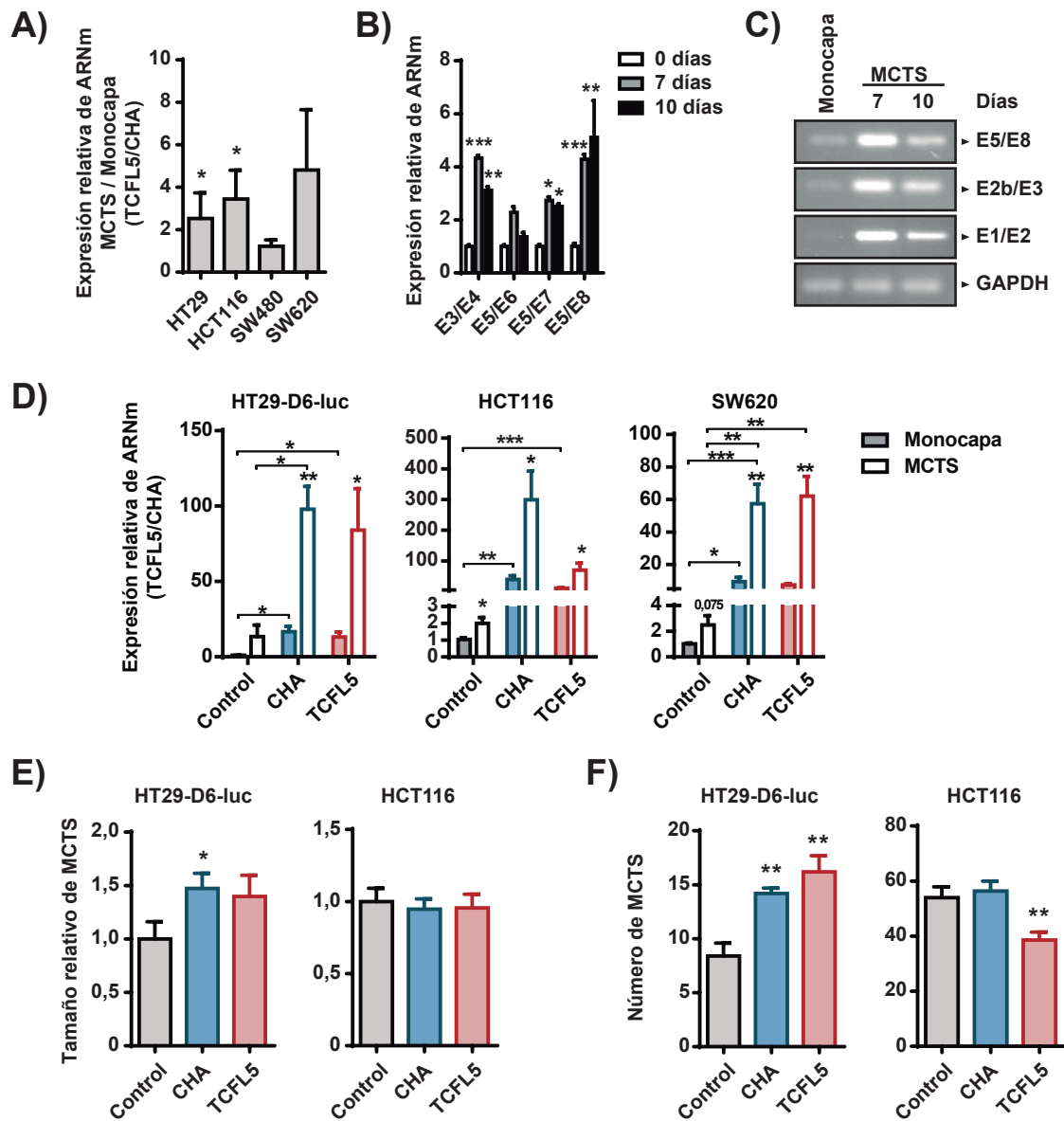


Figura 23. *TCFL5* y *CHA* se inducen en MCTS y alteran su formación de manera dependiente de línea celular. **A)** Análisis de la expresión por qPCR de *TCFL5/CHA* en la formación de MCTS de las líneas celulares de cáncer de colon HT29, HCT116, SW480 y SW620. **B)** Análisis de la expresión por qPCR de las uniones de exones E3/E4, E5/E6, E5/E7 y E5/E8 en la formación de MCTS a día 7 y 10 de la línea celular HCT116. **C)** Análisis de la expresión por PCR de *GAPDH* y de las uniones de exones E5/E8, E2b/E3 y E1/E2 en monocapa y MCTS a día 7 y 10 de la línea celular HCT116. **D)** Análisis de la expresión por qPCR de *TCFL5/CHA* en monocapa (barras rellenas) y en MCTS (barras sin rellenas) de las líneas celulares estables HT29-D6-luc, HCT116 y SW620 control (gris) y que sobreexpresan CHA (azul) y TCFL5 (rojo). **E)** Tamaño relativo de los MCTS a día 7 de las líneas estables HT29-D6-luc y HCT116 control (gris) y que sobreexpresan CHA (azul) y TCFL5 (rojo). **F)** Número de MCTS a día 7 de las líneas estables HT29-D6-luc y HCT116 control (gris) y que sobreexpresan CHA (azul) y TCFL5 (rojo). En las gráficas se representa la media \pm SEM de tres experimentos independientes y su estadística (* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$). Los asteriscos sobre la barra indican la estadística respecto al control. Los asteriscos sobre la línea indican la estadística entre las condiciones que une.

mayor expresión de todas las isoformas excepto de *TCFL5_E6*. Para profundizar y diferenciar entre las isoformas *TCFL5* y *CHA*, analizamos la expresión de las uniones de exones E5/E8, E2b/E3 (específica para *CHA*) y E1/E2 (que no reconoce la isoforma *CHA*) en la formación de MCTS de la línea celular HCT116 (Figura 23C). Las 3 uniones se indujeron durante la formación de MCTS confirmando la inducción de las isoformas *TCFL5* y *CHA*. En todos los casos el pico de expresión se dio a los 7 días de formación de MCTS. Estos datos confirman que durante la generación de MCTS hay una inducción de todas las isoformas de *TCFL5*, principalmente de *TCFL5* y *CHA*.

A continuación, quisimos estudiar el papel de las isoformas *TCFL5* y *CHA* en este proceso. Generamos MCTS de las líneas celulares estables HT29-D6-luc, HCT116 y SW620 con sobreexpresión de *TCFL5* y *CHA* observando nuevamente la inducción de *TCFL5/CHA* en MCTS respecto a monocapa en las 3 líneas control, aunque en la línea SW620 EV la expresión no fue estadísticamente significativa (Figura 23D). Además, en todas las líneas celulares con sobreexpresión de *TCFL5* y *CHA*, la expresión de estas isoformas en MCTS fue mayor que en las líneas EV, sugiriendo un efecto sinérgico entre la formación de MCTS y la sobreexpresión forzada de las isoformas. Por último, estudiamos los MCTS generados por estas líneas celulares. Se midió el tamaño relativo de estos (Figura 23E) y el número de MCTS (Figura 23F). En la línea HT29-D6-luc, observamos que la sobreexpresión de *CHA* produjo un aumento del tamaño y del número de MCTS mientras que *TCFL5* aumento el número de MCTS. En la línea HCT116, la transfección de *TCFL5* tuvo efecto reduciendo el número de MCTS, pero no el tamaño de estos. Además, se analizó la formación de MCTS en las líneas HCT116 *TCFL5*^{-/-}. En este caso, en las líneas KO23 y KO34 los MCTS no llegaron a generarse (datos no mostrados) por lo que se puede inferir que la deficiencia total de *TCFL5/CHA* inhibe la formación de MCTS indicando que la inducción de *TCFL5* y/o *CHA* es necesaria en la formación de MCTS.

3.2. *TCFL5* y *CHA* no producen cambios significativos en los marcadores de CSCs.

Dado que *TCFL5* y *CHA* tienen un papel fundamental en la formación de MCTS, y debido a que la formación de estos genera un aumento en la población de CSCs, quisimos estudiar la posible implicación de estas isoformas sobre esta población celular. Para ello analizamos por citometría de flujo la expresión de algunos de los marcadores más utilizados en líneas de cáncer de colon: Aldehído Deshidrogenasa I (ALDH1) [144], LGR5 [145], EPHb2 [146] o CD133 [147]. No se observaron diferencias significativas en ALDH1 con la expresión de ninguna de las isoformas, aunque existe una tendencia de mayor actividad en las líneas *TCFL5* en HCT116 (Figura 24A). En la línea celular HCT116, todas las líneas presentaron expresión de CD133 mientras que con *TCFL5* se redujo a un 60% el número de células positivas, aunque no fue estadísticamente significativo (Figura 24B). En cambio, en línea HT29-D6-luc, ni la sobreexpresión de *TCFL5* ni de *CHA* tuvieron efecto en la expresión de CD133. En el caso de EPhB2 y LGR5, la expresión en un cultivo en monocapa de células HCT116 fue muy baja. Tras la formación de esferoides de 7 días, hubo una inducción en la

expresión de EPHB2 en todas las líneas ligeramente mayor en las TCFL5 (Figura 24C). Para LGR5 tras la formación de MCTS hubo una inducción en la expresión de LGR5 en todas las líneas, aunque ligeramente menor para las células CHA y mayor en las TCFL5 (Figura 24D). Tomando en conjunto todos estos datos se observa una tendencia en la línea HCT116, principalmente por parte de TCFL5, de alterar la expresión de estos marcadores de CSCs aunque no de manera significativa.

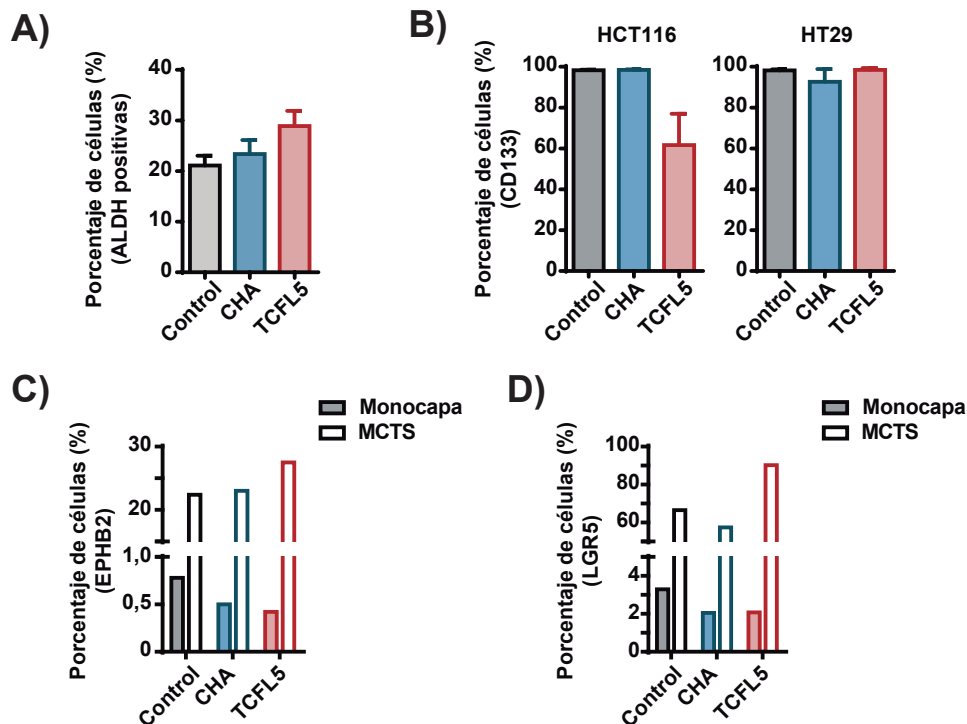


Figura 24. TCFL5 y CHA no alteran significativamente la expresión de los principales marcadores de CSCs en células de cáncer de colon. **A)** Análisis por citometría de flujo de la actividad del marcador ALDH1 en células estables HCT116 control (gris) y que sobreexpresan CHA (azul) y TCFL5 (rojo). **B)** Análisis de la expresión por citometría de flujo del marcador CD133 en células estables HT29 y HCT116 control (gris) y que sobreexpresan CHA (azul) y TCFL5 (rojo). **C-D)** Análisis de la expresión por citometría de flujo de los marcadores EPHB2 (C) LGR5 (D) en monocapa (barras rellenas) y en MCTS (barras sin relleno) de la línea celular estables HCT116 control (gris) y que sobreexpresan CHA (azul) y TCFL5 (rojo). En las gráficas se representa la media \pm SEM de tres experimentos independiente.

3.3. TCFL5 y CHA son represores de la expresión de genes de pluripotencia.

Además de marcadores propios de CSCs, quisimos estudiar si TCFL5 y/o CHA controlan el estado de pluripotencia de la célula estudiando los genes *SOX2*, *KLF4* y *NANOG*. Tras la formación de MCTS de las líneas celulares estables HT29-D6-luc y HCT116 que sobreexpresan TCFL5 y CHA, analizamos por qPCR la expresión de estos genes tanto en monocapa como en MCTS. En general, las células en monocapa no presentaron diferencias significativas debido a la expresión de CHA o TCFL5 (Figura 25). En la línea celular HT29-D6-luc la formación de MCTS indujo una fuerte expresión de *SOX2* y *KLF4*, pero mientras que la sobreexpresión de CHA inhibió la inducción de *SOX2*, la de TCFL5 no tuvo efecto. Al contrario, la inducción de *KLF4* no fue afectada por CHA, pero completamente inhibida por TCFL5 (Figura 25A). En la línea celular HCT116 (Figura 25B), no se observaron cambios significativos en la expresión de *SOX2* por la presencia de TCFL5 y CHA, aunque se apreció una tendencia a una mayor inducción en los MCTS formados en presencia de

CHA. En el caso de la expresión de *KLF4*, la sobreexpresión tanto de *TCFL5* como de *CHA* produjo una disminución de los niveles basales de este gen. La formación de MCTS no produjo efectos en su expresión. También analizamos la expresión de *NANOG* cuya inducción producida por la formación de MCTS se vio ligeramente reducida con la presencia de *CHA* y *TCFL5*.

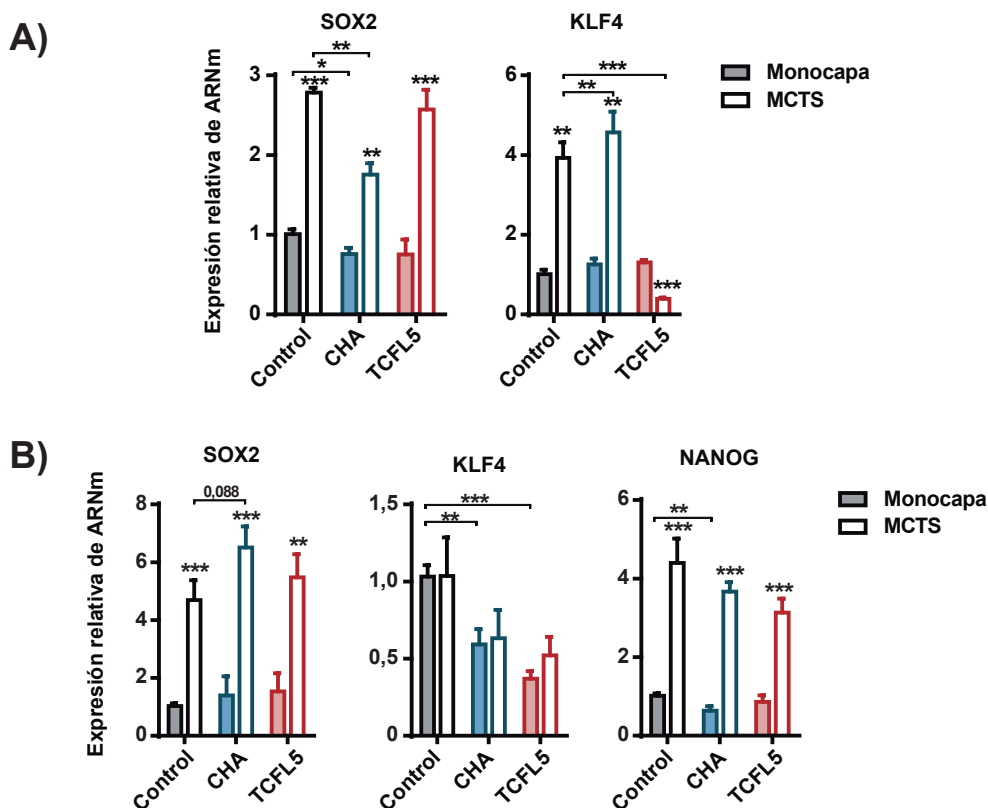


Figura 25. *TCFL5* y *CHA* alteran la expresión de marcadores de pluripotencia en la formación de MCTS. Análisis por qPCR de la expresión de genes de pluripotencia en monocapa (barras rellenas) y MCTS a 7 días (barras sin relleno) en las líneas celulares estables HT29-D6-luc y HCT116 control (gris) y que sobreexpresan *CHA* (azul) y *TCFL5* (rojo). **A)** Expresión de *SOX2* y *KLF4* en las líneas HT29-D6-luc. **B)** Expresión de *SOX2*, *KLF4* y *NANOG* en las líneas HCT116. En las gráficas se representa la media \pm SEM de tres experimentos independientes y su estadística (* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$). Los asteriscos sobre la barra indican la estadística respecto a su expresión en monocapa. Los asteriscos sobre la línea indican la estadística entre las condiciones que une.

Como hemos visto anteriormente, *TCFL5* y *CHA* tienen funciones diferentes y se regulan mutuamente induciendo la expresión el uno del otro. Esto puede provocar una distorsión de los resultados y los efectos observados en MCTS no ser causa de la sobreexpresión exógena de estas sino de un efecto secundario por parte de su regulación recíproca. Por ello quisimos descartar que los efectos observados en la expresión de marcadores de pluripotencia fuesen debidos a este motivo. En células HCT116 *TCFL5*^{-/-} (KO23 y KO34) reconstituimos la expresión de cada isoforma para observar el efecto único de cada una de ellas. Por qPCR analizamos la expresión de los genes *TCFL5*/*CHA*, para comprobar la recuperación de la expresión, *SOX2*, *NANOG*, *KLF4* y *cMYC* (Figura 26A). La sobreexpresión condujo a mayores niveles de *TCFL5*/*CHA* en las células HCT116 *TCFL5*^{-/-} que en la línea WT6. Con respecto a la expresión de *SOX2* y *NANOG*, la línea control WT6 presentó menor expresión con la sobreexpresión de *CHA* y mayor expresión con *TCFL5*. Sin embargo, las líneas KO23 y KO34 presentaron mayores niveles de expresión de *SOX2* y *NANOG* que se vieron revertidos con la sobreexpresión de *TCFL5* y *CHA*. En el caso de *KLF4*, su expresión en la línea

WT6 fue menor con TCFL5 y mayor con CHA mientras que las líneas KO23 y KO34 presentaron menores niveles de expresión que la línea WT6 que no se vieron modificados tras la sobreexpresión de TCFL5 y CHA. Respecto a la expresión de *cMYC*, únicamente la sobreexpresión de TCFL5 y CHA en las líneas KO23 y KO34 produjo una disminución en su expresión. Estos resultados sugieren un papel represor de TCFL5 y CHA en la expresión de *SOX2* y *NANOG* mientras que para el caso de *KLF4*, la situación es más compleja pues su deficiencia completa inhibiría su expresión que no es recuperable ni por TCFL5 ni por CHA, aunque la sobreexpresión de TCFL5 en WT podría inhibirlo. Para comprobar a nivel de proteína el efecto de TCFL5 y CHA sobre *SOX2* y *NANOG*, analizamos su expresión mediante WB (Figura 26B). Tanto la expresión de CHA como de TCFL5 sólo fue detectable tras la sobreexpresión transitoria. El análisis densitométrico de las bandas reveló que la expresión de *SOX2* y *NANOG* fue mayor en las líneas KO23 y KO34 respecto a la WT6 y HCT116, en concordancia con los niveles de ARNm. La sobreexpresión transitoria de CHA no tuvo efecto en *SOX2* y *NANOG* en las líneas HCT116 y control WT6, mientras que disminuyó la expresión

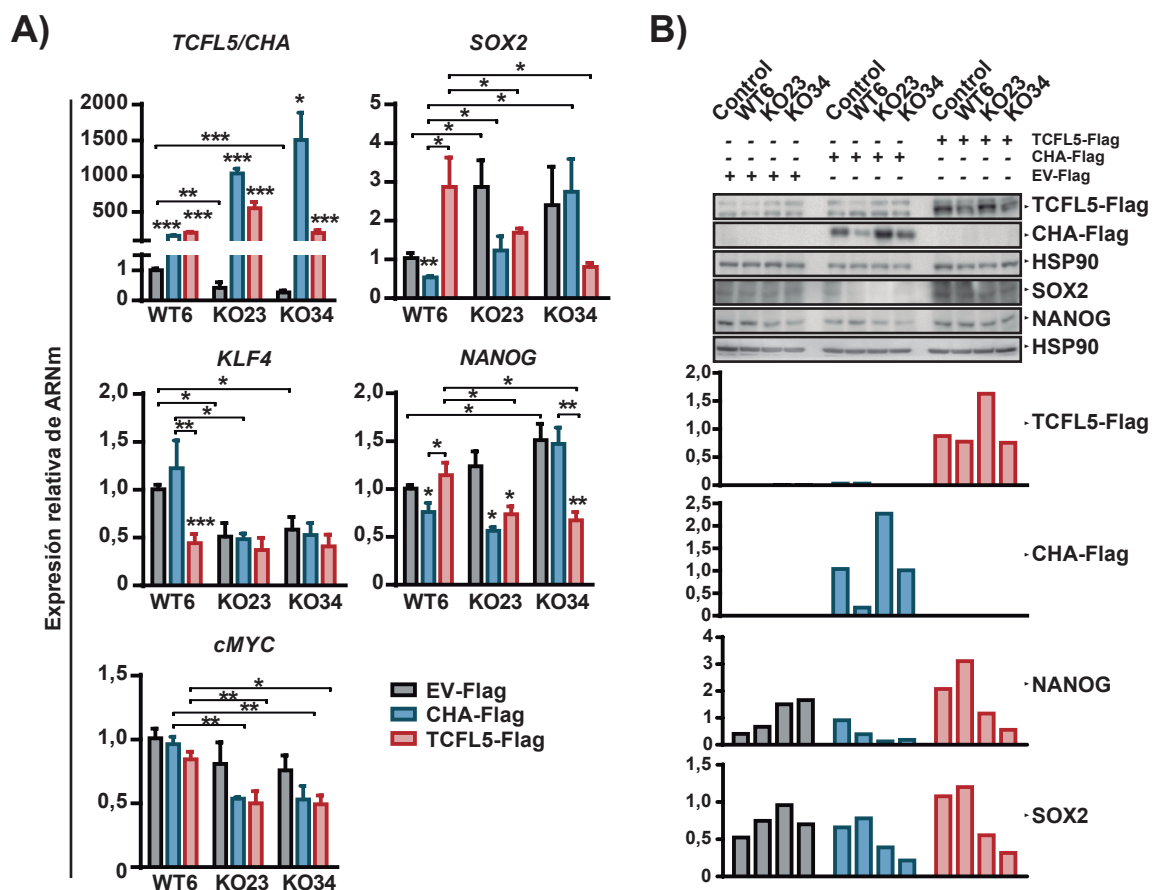


Figura 26. TCFL5 y CHA modulan la expresión de marcadores de pluripotencia. Análisis de la expresión de los marcadores de pluripotencia en células HCT116 TCFL5^{+/+} clon WT6 y HCT116 TCFL5^{-/-} clon KO23 y clon KO34 tras la reconstitución de la expresión de TCFL5 y CHA con la transfección de los plásmidos Control-Flag (gris), CHA-Flag (azul) y TCFL5-Flag (rojo). **A)** Análisis por qPCR de la expresión de *TCFL5/CHA*, *SOX2*, *KLF4*, *NANOG* y *cMYC*. **B)** Arriba, análisis de la expresión por Western-blot de la proteína de TCFL5, CHA, NANOG, SOX2 y HSP90 como control de carga. Abajo, cuantificación por densitometría del WB normalizando con la expresión de HSP90. En las gráficas se representa la media \pm SEM de tres experimentos independientes y su estadística (* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$). Los asteriscos sobre la barra indican la estadística respecto a su expresión control. Los asteriscos sobre la línea indican la estadística entre las condiciones que une.

aumentada de las líneas KO23 y KO34 en concordancia con los niveles de ARNm. La expresión transitoria de TCFL5 indujo SOX2 y NANOG en las líneas HCT116 y WT6 mientras que asimismo revirtió la inducción observada por la deficiencia. Estos resultados evidencian el papel represor de TCFL5 y CHA sobre SOX2 y NANOG e indican que, en células tumorales no manipuladas genéticamente, TCFL5 presenta una función distinta sobre estos genes implicando algún tipo de regulación adicional.

3.4. TCFL5 y CHA modifican la expresión génica en los tumores xenoinjertados.

Quisimos ampliar nuestro estudio sobre la expresión génica de los marcadores de pluripotencia analizando su expresión por qPCR a día final del experimento en los tumores xenoinjertados generados por las líneas estables HT29-D6-luc y HCT116 que sobreexpresan TCFL5 y CHA, y por las líneas HCT116 TCFL5^{-/-}. En los tumores generados por la línea HT29-D6-luc (Figura 27A), la sobreexpresión de *TCFL5* y *CHA* se perdió en algunos tumores a lo largo del experimento por lo que a día final no se observaron diferencias de expresión de estas isoformas con respecto al control. Los tumores con sobreexpresión de CHA, que presentaron un menor crecimiento (Ver Figura 15) presentaron mayores niveles de *SOX2*, *KLF4* y *NANOG*, aunque no significativo para estos dos últimos. Los tumores TCFL5, que crecieron ligeramente más despacio y al final tenían el mismo tamaño, únicamente presentaron mayor expresión de *NANOG* sin verse afectado el resto de genes. En los tumores generados por la línea HCT116 (Figura 27B), si se observó que la sobreexpresión de *TCFL5* y *CHA* se mantuvo en los tumores a día final del experimento. Los tumores con sobreexpresión de CHA, que este caso, tuvieron un crecimiento similar a los tumores control (ver Figura 15) presentaron una menor expresión de *SOX2* mientras que aumentó la expresión de *KLF4*. Por su lado, TCFL5 que en esta línea si generó tumores menos agresivos, tuvieron mayor expresión de *SOX2* y *TP53*.

Finalmente, en los tumores generados por las líneas HCT116 TCFL5^{-/-} que crecían mucho menos que los controles (ver Figura 15) se mantuvo la deficiencia en *TCFL5* y *CHA* en los tumores de las líneas KO23 y KO34 como era de esperar (Figura 27C). La expresión de *SOX2* y *NANOG* se vio aumentada sólo en los tumores de la línea KO34. La expresión de *KLF4* fue menor en los tumores de las líneas KO23 y KO34 con respecto al control en concordancia con lo observado en las propias líneas celulares (ver Figura 27). Por último, la expresión de *TP53* también se vio reducida por la deficiencia de TCFL5 y CHA.

3.5. TCFL5 reprime la actividad de SOX2, NANOG y KLF4.

Además de regular la expresión de SOX2, NANOG y KLF4, quisimos estudiar si TCFL5 y CHA alteran su actividad transcripcional. Se realizaron ensayos de actividad luciferasa en la que esta se encuentra bajo el control de un promotor mínimo para una expresión basal y varios sitios de

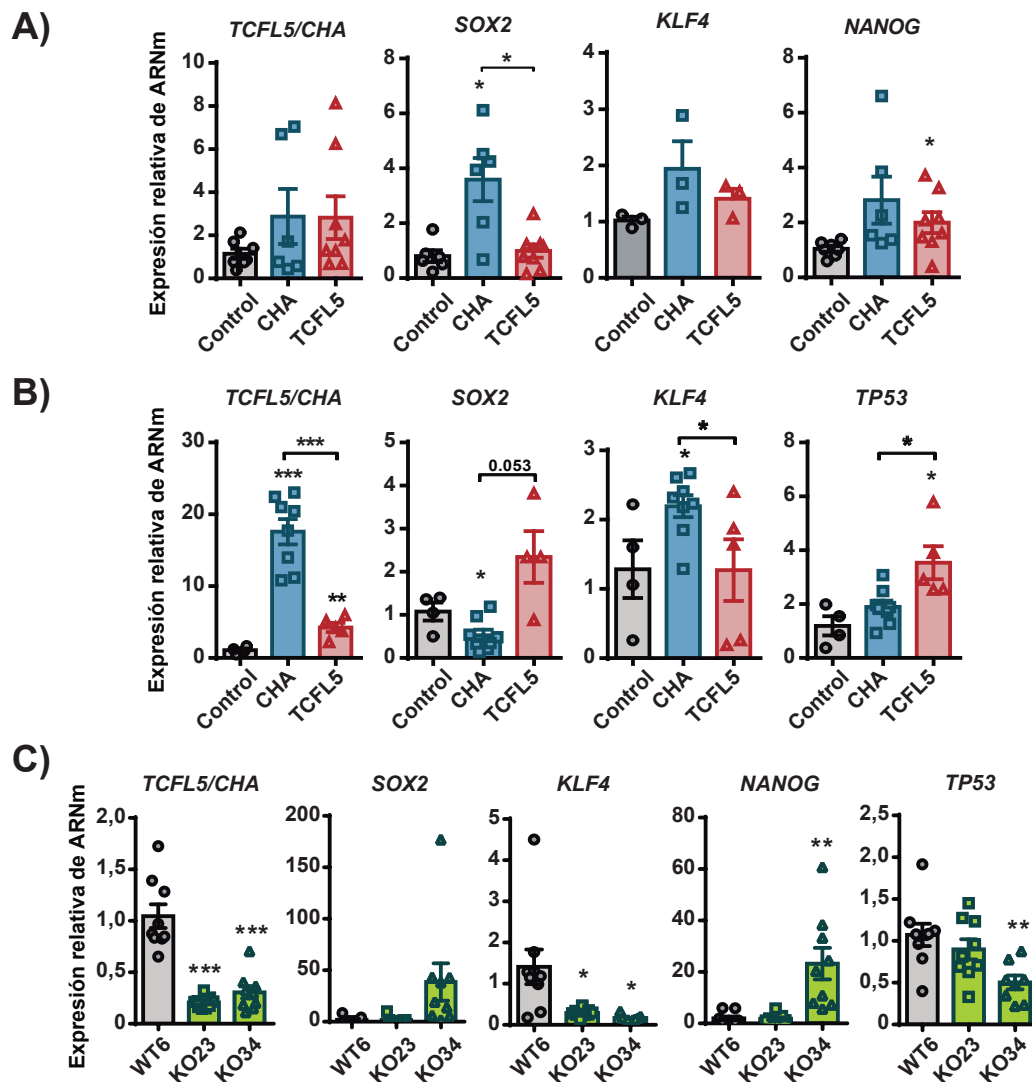


Figura 27. La sobreexpresión y deficiencia de TCFL5 y CHA modula la expresión de marcadores de pluripotencia en los tumores xenoinjertados. **A)** Análisis de la expresión por qPCR de *TCFL5/CHA*, *SOX2*, *KLF4* y *NANOG* en los tumores xenoinjertados de las líneas celulares HT29-D6-luc control (gris) y que sobreexpresan CHA (azul) y TCFL5 (rojo). **B)** Análisis de la expresión por qPCR de *TCFL5/CHA*, *SOX2*, *KLF4* y *TP53* en los tumores xenoinjertados de las líneas celulares HCT116 control (gris) y que sobreexpresan CHA (azul) y TCFL5 (rojo). **C)** Análisis de la expresión de *TCFL5/CHA*, *SOX2*, *KLF4*, *NANOG* y *TP53* en los tumores xenoinjertados de las líneas celulares HCT116 TCFL5^{+/+} clon WT6 (gris) y HCT116 TCFL5^{-/-} clon KO23 (verde) y clon KO34 (verde). En las gráficas se representa la expresión de cada tumor analizado, así como la media \pm SEM y su estadística (*p<0,05; **p<0,01; ***p<0,001). Los asteriscos sobre la barra indican la estadística respecto a la expresión de la condición control. Los asteriscos sobre la línea indican la estadística entre las condiciones que une.

respuesta a SOX2 (6 sitios), NANOG (2 sitios) y KLF4 (5 sitios) para determinar la actividad de cada una de estas proteínas (Figura 28A). Se transfectaron las líneas HCT116 TCFL5^{-/-} con las isoformas TCFL5 y CHA para reconstituir la expresión específica de dichas isoformas. Además, como control positivo, se utilizó el plásmido pEP4 que expresa las proteínas SOX2, KLF4 y NANOG entre otras [148]. En cuanto a la actividad de SOX2 (Figura 28B), la sobreexpresión de TCFL5 produjo una reducción en la actividad mientras que no se encontraron diferencias con la sobreexpresión de CHA en la línea control WT6. Las líneas KO23 y KO34 no presentaron diferencias con respecto a la línea WT6 y la recuperación de la expresión de TCFL5 y CHA tampoco produjo cambios significativos en la actividad de SOX2. No obstante, la transfección con el plásmido control positivo pEP4 no produjo

grandes variaciones en la actividad en células control por lo que es probable que este elemento de respuesta no funcionase de forma correcta en la línea HCT116. Con respecto a la actividad de NANOG (Figura 28C), el control positivo pEP4 produjo un aumento muy importante de su actividad en todos los casos. En la línea WT6, la sobreexpresión de CHA y TCFL5 produjo un aumento y una disminución de la actividad NANOG, respectivamente. En las líneas KO23 y KO34, la actividad de NANOG no se vio afectada por la deficiencia, sin embargo, la reexpresión de TCFL5 produjo una disminución de la actividad mientras que CHA no tuvo efecto. La actividad de KLF4 también se vio inducida con el plásmido control pEP4 (Figura 28D). En la línea WT6, la sobreexpresión de CHA produjo un aumento de la actividad de KLF4 mientras que TCFL5 produjo una disminución, aunque no estadísticamente significativa en ningún caso. En las líneas KO23 y KO34, no se observaron cambios en la actividad basal de KLF4, pero si tendió a aumentar con la expresión de CHA y disminuyó con la transfección de TCFL5. En conjunto, estos datos sugieren que TCFL5, pero no de CHA, disminuyen la actividad de estas proteínas.

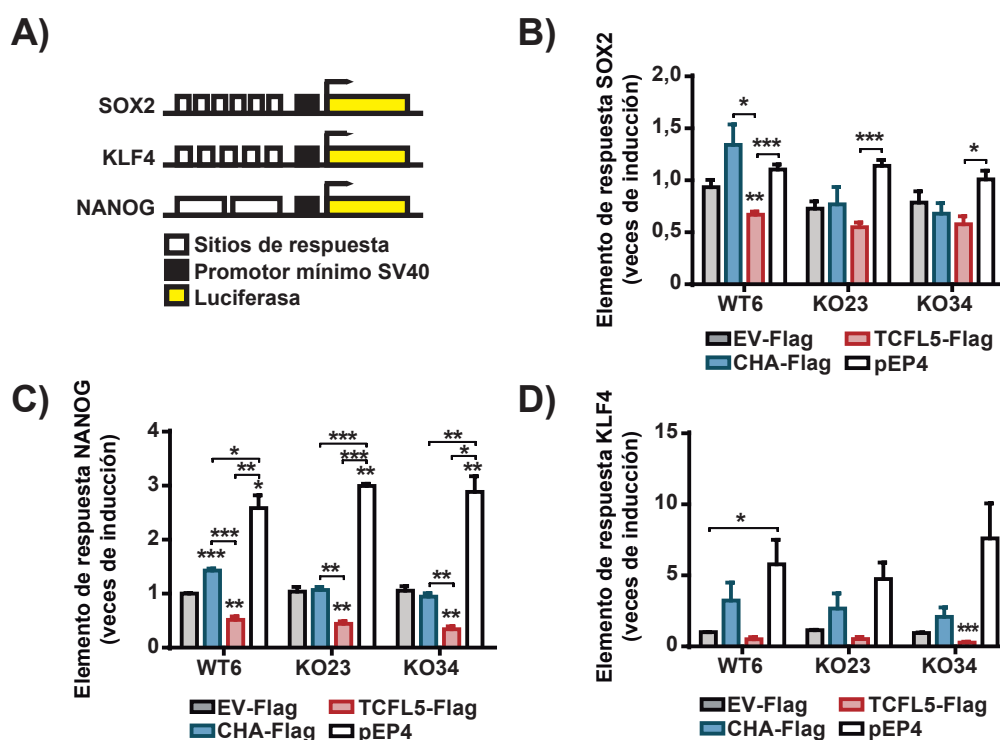


Figura 28. TCFL5 reprime la actividad de SOX2, KLF4 y NANOG. Análisis de la actividad por luminiscencia de SOX2, KLF4 y NANOG en células HCT116 TCFL5^{+/+} clon WT6 y HCT116 TCFL5^{-/-} clon KO23 y clon KO34 tras la reconstitución de la expresión de TCFL5 y CHA con la transfección de los plásmidos Control-Flag (gris), CHA-Flag (azul), TCFL5-Flag (rojo) y pEP4 (blanco) como control positivo. **A)** Esquema de los elementos de repuestas para SOX2, KLF4 y NANOG utilizados. **B)** Actividad del elemento de respuesta para SOX2. **C)** Actividad del elemento de respuesta para NANOG. **D)** Actividad del elemento de respuesta para KLF4. En las gráficas se representa la media \pm SEM de tres experimentos independientes y su estadística (* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$). Los asteriscos sobre la barra indican la estadística respecto a la expresión de la condición control. Los asteriscos sobre la línea indican la estadística entre las condiciones que une.

3.6. TCFL5 y CHA se unen al promotor de SOX2 y KLF4.

Una vez comprobado que principalmente TCFL5 modifica la expresión y actividad de los marcadores de pluripotencia, quisimos estudiar si este efecto se debía a una regulación

transcripcional directa de estos genes por parte de TCFL5. Para ello realizamos experimentos de inmunoprecipitación de cromatina transfectando las células HCT116 con las isoformas TCFL5 y CHA fusionadas a la etiqueta FLAG para la inmunoprecipitación específica de cada isoforma. Tras la inmunoprecipitación, se amplificaron por qPCR las regiones específicas del promotor de *SOX2* y *KLF4* para determinar la unión de TCFL5 y CHA. En el caso de *SOX2*, se analizaron dos regiones del promotor definidas previamente que controlan la expresión de este gen: la región CORE, más próxima al inicio de transcripción, y la región *enhancer* o R1, más alejada del inicio de transcripción [149]. Se estudiaron dos regiones de la zona CORE (CORE_01 y CORE_02), una región de la zona R1 (R1_01) y una zona alejada de este R1 transcripcionalmente inactiva como control (R1_02) (Figura 29A). Tanto TCFL5 como CHA se unieron a las regiones CORE_01 y CORE_02. Sin embargo, aunque se observó unión en la región R1_01 fue de igual magnitud a la observada para la zona de control negativo R1_02 por lo que lo consideramos una amplificación residual no específica. Estos datos indican que TCFL5 y CHA se unen a la región CORE del promotor de *SOX2*, pero no a la región *enhancer*. En el caso de *KLF4*, se estudiaron dos regiones del promotor: una región próxima el inicio de transcripción (positiva) y una región alejada y transcripcionalmente no activa (negativa) (Figura 29B). Tanto TCFL5 como CHA amplificaron la región positiva del promotor de *KLF4*, pero no la negativa indicando que ambos factores de transcripción se unen al promotor de *KLF4*.

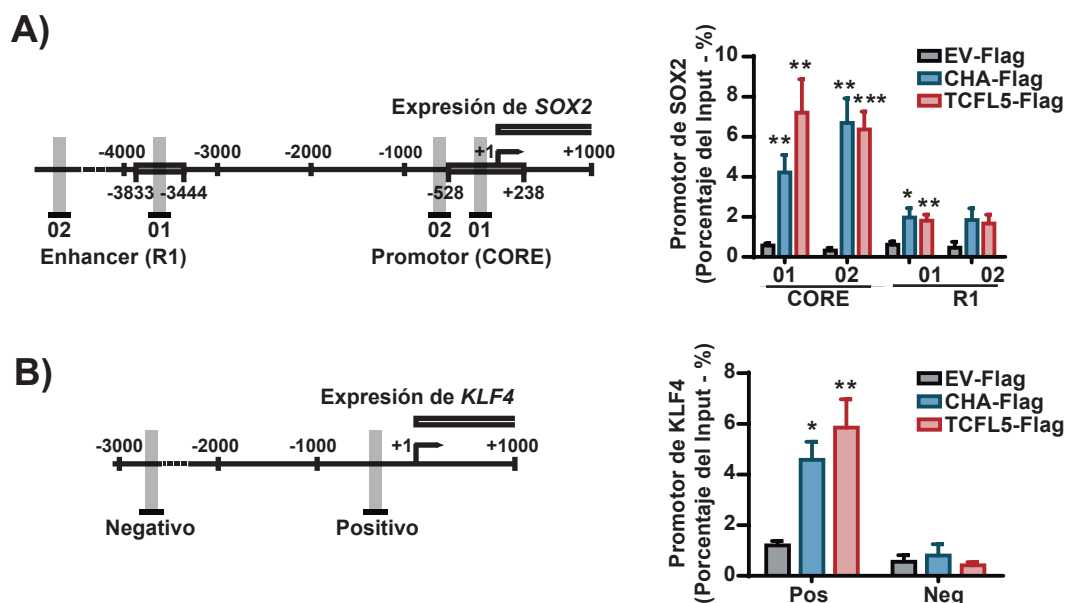


Figura 29. TCFL5 y CHA se unen a los promotores de *SOX2* y *KLF4*. Análisis por ensayos de CHIP de la unión de TCFL5 y CHA a los promotores de *SOX2* y *KLF4* en la línea celular HCT116 transflectada con los plásmidos Control-Flag (gris), CHA-Flag (azul) y TCFL5-Flag (rojo). **A)** A la izquierda, esquema de las regiones CORE y R1 del promotor de *SOX2* y las regiones estudiadas en los experimentos de CHIP. A la derecha, análisis por qPCR de la unión de TCFL5 y CHA al promotor de *SOX2*. **B)** A la izquierda, esquema del promotor de *KLF4* y las estudiadas en los experimentos de CHIP. A la derecha, análisis por qPCR de la unión de TCFL5 y CHA al promotor de *KLF4*. En las gráficas se representa la media \pm SEM de tres experimentos independientes y su estadística respecto a la transfección EV-Flag de cada región estudiada (* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$).

3.7. TCFL5 y CHA reprimen la actividad del promotor de *SOX2*.

Analizamos la función de TCFL5 y CHA sobre las regiones CORE y R1 del promotor de *SOX2* para determinar si la unión al promotor por parte de estos factores de transcripción se traduce

en una regulación de la actividad de este. Se utilizaron ensayos de actividad luciferasa con plásmidos donde esta actividad se encuentra bajo el control de las regiones CORE y R1 del promotor de *SOX2* [149] (Figura 30A). Se transfectaron en la línea celular HCT116 y HEK293 las isoformas TCFL5 y CHA y el plásmido control positivo pEP4. En el caso de la regulación de la región CORE (Figura 30B), la transfección con el plásmido pEP4 produjo un aumento en la actividad del promotor tanto en líneas HCT116 como en HEK293. Sin embargo, la sobreexpresión de TCFL5 o CHA no tuvo efecto en ninguna de las dos líneas. En la regulación de la región R1 (Figura 30C) la transfección con el plásmido pEP4 tuvo una cierta tendencia a una mayor actividad de este promotor, aunque no fue significativa en ninguna de las líneas. En la línea HEK293, la sobreexpresión de TCFL5 aumentó la actividad del promotor R1 y aun en mayor cantidad junto a pEP4. De nuevo, para evitar interferencias debidas a la regulación recíproca de las isoformas, realizamos los mismos ensayos de actividad luciferasa en células HCT116 TCFL5^{-/-} tras la transfección de las isoformas TCFL5 y CHA para reconstituir la expresión de estas. La actividad de la región CORE (Figura 30D), en la

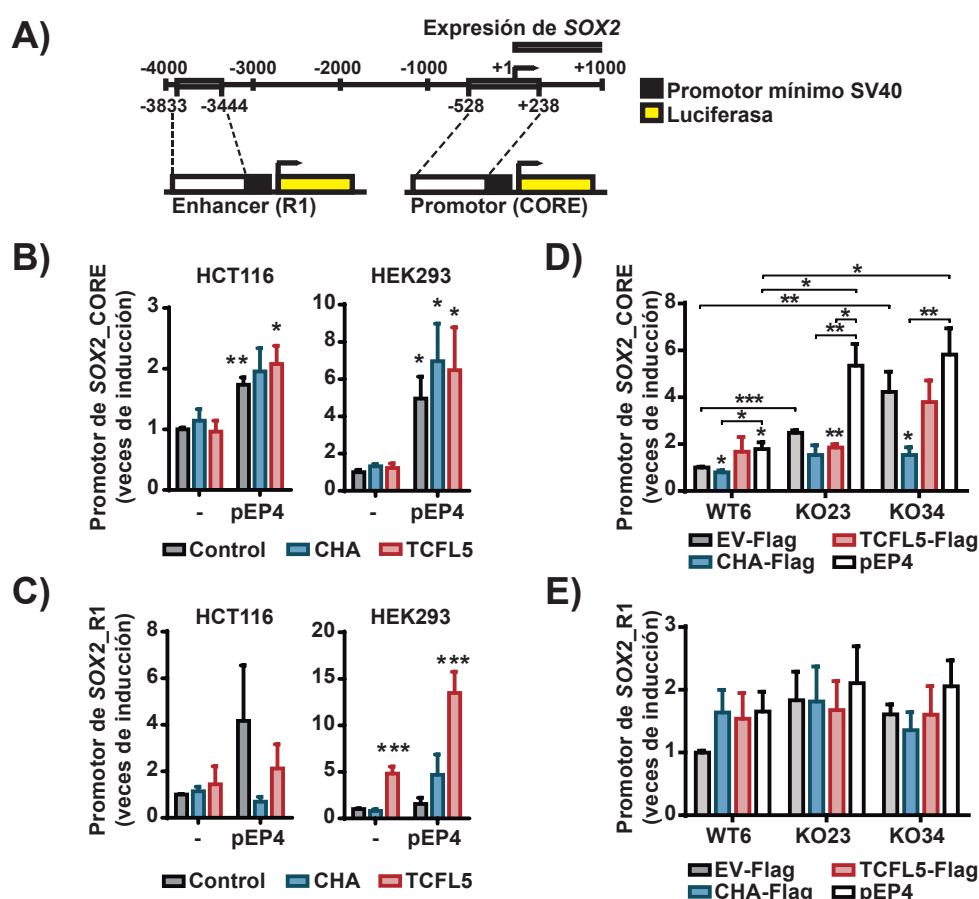


Figura 30. TCFL5 y CHA regulan la actividad del promotor de *SOX2*. Análisis de la activación del promotor de *SOX2* por TCFL5 y CHA mediante ensayos de luminiscencia. **A)** Esquema de las regiones del promotor de *SOX2* estudiadas y de los plásmidos *SOX2_CORE* y *SOX2_R1* utilizados. Esquema adaptado de *Leis O, et al 2012*. **B-C)** Actividad del promotor *SOX2_CORE* (B) y *SOX2_R1* (C) en las líneas HCT116 y HEK293 sobreexpresadas transitoriamente con plásmido Control-Flag (gris), CHA-Flag (azul), TCFL5-Flag (rojo) y pEP4. **D-E)** Actividad del promotor *SOX2_CORE* (D) y *SOX2_R1* (E) en células HCT116 TCFL5^{+/+} clon WT6 y HCT116 TCFL5^{-/-} clon KO23 y clon KO34 tras la reconstitución de la expresión de TCFL5 y CHA con la transfección de los plásmidos Control-Flag (gris), CHA-Flag (azul), TCFL5-Flag (rojo) y pEP4 (blanco) como control positivo. En las gráficas se representa la media \pm SEM de tres experimentos independientes y su estadística (* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$). Los asteriscos sobre la barra indican la estadística respecto a la expresión de la condición control. Los asteriscos sobre la línea indican la estadística entre las condiciones que une.

línea WT6, se vio disminuida con la sobreexpresión de CHA mientras que la transfección con el plásmido pEP4 la aumentó. Las líneas KO23 y KO34 presentaron mayor actividad basal de la región promotora CORE de *SOX2* que la línea WT6 y, además, esta mayor actividad se vio incrementada tras la transfección con el plásmido pEP4. La re-expresión de *TCFL5* o de CHA en estas líneas KO23 y KO34 redujo el incremento basal de la actividad de esta región promotora. Por último, la actividad de la región promotora R1 no se vio significativamente afectada por ninguna de las condiciones estudiadas (Figura 30E). Estos datos junto con los anteriores indican que, efectivamente, existe una regulación de *SOX2* por parte de *TCFL5* y CHA uniéndose a su promotor y reprimiendo su transcripción sugiriendo un posible papel represor en procesos de pluripotencia.

4. Expresión de *Tcf15* en etapas tempranas del desarrollo embrionario de ratón y su papel en diferenciación.

La pluripotencia es la capacidad de las células embrionarias por la cual pueden generar todos los tipos de células somáticas, así como las células de la línea germinal de un adulto. Esta capacidad es transitoria y se produce solo durante las primeras etapas del desarrollo embrionario [74]. Como ya hemos visto, *TCFL5* es un factor de transcripción poco expresado en humanos y ratón exceptuando testículo y cerebro donde presenta una mayor expresión. Sin embargo, su control transcripcional sobre genes implicados en el mantenimiento de la pluripotencia como son *SOX2*, *KLF4* y *NANOG*, nos planteó la posibilidad de un posible papel durante los estadios tempranos del desarrollo embrionario controlando la expresión de dichos genes.

4.1. *Tcf15* se expresa en etapas tempranas del desarrollo embrionario y en células embrionarias de ratón.

Quisimos estudiar la expresión de las isoformas de *TCFL5* durante las primeras etapas del desarrollo embrionario de humano. Realizamos una búsqueda en bases de datos y obtuvimos los datos de microarrays de expresión en el que se estudió la expresión génica durante las primeras etapas del desarrollo de humanos y ratón (1 célula, 2 células, 4 células, 8 células, mórula y blástula) [64]. En humanos, la expresión de la isoforma *TCFL5_E6* se encontró elevada en 1 célula y fue disminuyendo progresivamente durante las distintas etapas del desarrollo hasta permanecer en niveles bajos en el estadio de blástula (Figura 31A). Sin embargo, la expresión de *TCFL5/CHA* fue baja en el estadio de 1 célula y aumentó progresivamente hasta blástula donde presenta su máxima expresión. El patrón de expresión de estas isoformas es completamente opuesto y sugiere un posible papel diferencial durante este proceso. Paralelamente, analizamos la expresión de *Tcf15* durante el desarrollo embrionario de ratón [64] (Figura 31B). En este caso el microarray no permite diferenciar entre isoformas, pero se observó que *Tcf15* presentó una baja expresión durante los primeros estadios y aumentó progresivamente durante el proceso de desarrollo embrionario al igual que sucede en humanos. Estos resultados indican que hay una fuerte y variable expresión de *TCFL5* tanto en

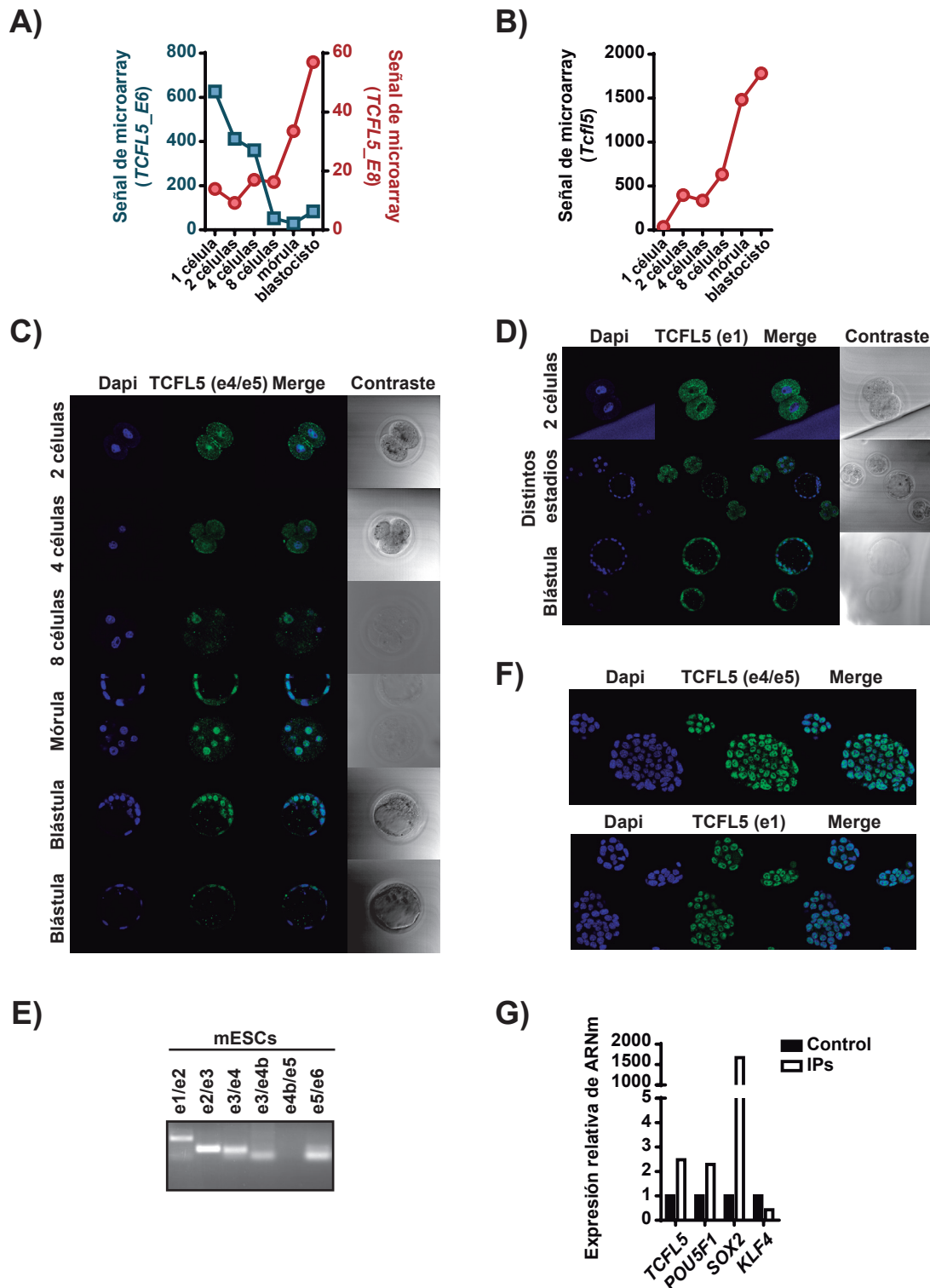


Figura 31. TCFL5 se encuentra expresado durante el periodo de preimplantación del desarrollo embrionario de ratón y en células pluripotentes. **A)** Análisis de la expresión en microarrays de las isoformas *TCFL5_E8/CHA* (rojo) y *TCFL5_E6* (azul) durante el periodo de preimplantación del desarrollo embrionario humano. **B)** Análisis de la expresión por microarray de *Tcf5* durante el desarrollo embrionario de ratón. **C-D)** Análisis de la expresión de TCFL5 por inmunofluorescencia durante las distintas fases del periodo de preimplantación del desarrollo embrionario del ratón. Se realizó un marcaje nuclear con Dapi (azul) y de TCFL5 con dos anticuerpos: TCFL5 e4/e5 (figura C, verde) y TCFL5 e1 (figura D, verde). **E)** Análisis de la expresión de las uniones de exones e1/e2, e2/e3, e3/e4, e3/e4b, e4b/e5 y e5/e6 de *Tcf5* en células embrionarias de ratón. **F)** Análisis de la expresión de TCFL5 en células embrionaria de ratón por inmunofluorescencia. Dapi (azul) y TCFL5 (verde). **G)** Análisis de la expresión por qPCR de *TCFL5/CHA*, *POU5F1*, *SOX2* y *KLF4* tras la formación de células IPs a partir de fibroblastos humanos.

humanos como en ratón durante el desarrollo embrionario y por tanto puede tener un papel en este proceso.

Con el objetivo de confirmar su expresión en el desarrollo embrionario, realizamos inmunofluorescencias para la detección de TCFL5 en los distintos estadios del desarrollo embrionario de ratón. Utilizamos dos anticuerpos distintos para detectar la expresión de TCFL5: uno de ellos para la región bHLH en los exones e4 y e5 (α -TCFL5) y otro para la región del exón e1 (α -E1). Con el primer anticuerpo (Figura 31C), en el primer estadio de 2 células, se observó expresión principalmente nuclear, aunque también citoplasmática. En el estadio de 4 células la expresión de TCFL5 fue también ubicua en núcleo y citoplasma, aunque la expresión nuclear fue menor. A partir de 8 células y mórula, la expresión dejó de ser ubicua y se concentró principalmente en el núcleo. Finalmente, en el estadio de blástula, la expresión es principalmente nuclear y mayor en la ICM. Para el marcaje del anticuerpo α -E1 no pudimos obtener muestras de todas las etapas del desarrollo embrionario, sin embargo, el patrón de expresión fue distinto (Figura 31D). En el estadio de 2 células la expresión fue totalmente citoplasmática sin observarse marcaje nuclear en ningún caso. En el resto de estadios la expresión se encontró tanto en el núcleo como en el citoplasma para llegar a ser específica del núcleo en el estadio de blástula. Estos datos indican una expresión de TCFL5 durante el desarrollo embrionario de ratón y que cambia su localización durante este proceso. Además, el diferente marcaje con dos anticuerpos que reconocen las mismas isoformas implica una regulación alternativa en la expresión del primer exón como podía ser un procesamiento proteolítico.

Para continuar con el análisis de expresión de *Tcfl5*, estudiamos las isoformas expresadas en células embrionarias de ratón (mESC) que derivan de la ICM a través de la amplificación de las uniones de exones por PCR (Figura 31E). Las uniones de exones e1/e2, e2/e3, e3/e4, e3/e4b y e5/e6 se encontraron expresadas en estas células mientras que la unión de exones e4b/e5 no se detectó indicando que las células embrionarias de ratón expresan todas las isoformas de *Tcfl5*. Además, se realizaron inmunofluorescencias de estas células utilizando los anticuerpos para TCFL5 descritos anteriormente (Figura 31F). La expresión fue la misma para ambas tinciones encontrándose en el núcleo de estas células. Por último, quisimos estudiar la expresión de *TCFL5/CHA* en otro tipo de célula pluripotente como son las células IPs generadas a partir de fibroblastos humanos (Figura 31G). En las células IPs hubo una mayor expresión de *TCFL5/CHA* que en fibroblastos al igual que con los genes de pluripotencia *POU5F1* y *SOX2*. Sin embargo, la expresión de *KLF4* se vio reducida en células IPs.

4.2. La deficiencia en TCFL5 modifica la expresión de SOX2 y NANOG en células embrionaria de ratón.

Después de estudiar la expresión de TCFL5 durante el desarrollo embrionario y en células pluripotentes embrionarias de ratón generamos, en estas células, diferentes líneas deficientes en la

expresión de las isoformas de TCFL5 mediante la técnica *self cloning* CRISPR-Cas9 (Figura 32A). Las líneas e3-1 y e3-2 presentan una delección en el exón e3 por lo que estas células son deficientes para todas las isoformas de *Tcfl5* (*Tcfl5_e6*, *Tcfl5_e6b*, *Tcfl5_e6Δ5*, *Tcfl5_e4l* y *Tcfl5_e5*). La línea e4b tiene una delección en el exón e4b por lo que únicamente no expresa la isoforma *Tcfl5_e4l*. La línea e5 presenta una delección en el exón e5 lo que produce una deficiencia de las isoformas *Tcfl5_e6*, *Tcfl5_e6b* y *Tcfl5_e5*. Por último, la línea e6 tiene la delección en el exón e6 y no presenta expresión de las isoformas *Tcfl5_e6*, *Tcfl5_e6b* y *Tcfl5_e6Δ5*. El ADN de estas células fue secuenciado para comprobar la correcta delección (datos no mostrados). Por qPCR analizamos la expresión de *Tcfl5* amplificando diferentes uniones de exones para comprobar su expresión residual (Figura 32B). Observamos que las líneas e3-1 y e3-2, deficientes para todas las isoformas, no expresan las uniones de exones e3/e4 y e5/e6. Sin embargo, mientras que la línea e3-1 es también deficiente para la expresión de la unión e4/e4l, la línea e3-2 no lo es. La línea e4l, exclusiva para la deficiencia de la isoforma *Tcfl5_e4l*, no expresa ninguna unión de exones analizadas, sugiriendo la implicación de esta isoforma en la regulación del resto. Las líneas e5 y e6, exclusivas para ciertas isoformas, solo presentan deficiencia de la unión e5/e6 mientras que el resto se siguen expresando. Además, se analizó la expresión por qPCR de los marcadores de pluripotencia *Sox2* y *Nanog* en todas las líneas generadas (Figura 32C). Ninguna de las líneas deficientes en la expresión de TCFL5 generadas produjo cambios significativos en la expresión de estos genes. Por último, analizamos la expresión a nivel de proteína de SOX2 y

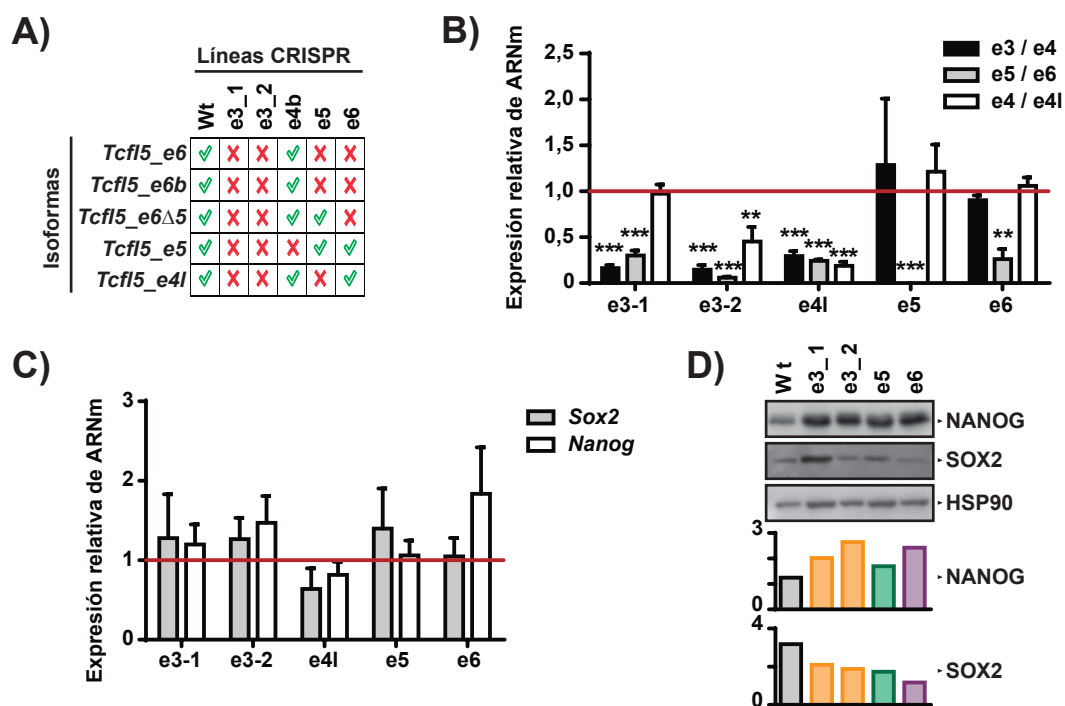


Figura 32. Generación de líneas embrionarias de ratón deficientes para la expresión de TCFL5. **A)** Cuadro resumen de la expresión de las isoformas de *Tcfl5* en las líneas embrionaria de ratón generadas por *self-cloning* CRISPR-Cas9 *Tcfl5*^{-/-} e3-1, e3-2, e4l, e5 y e6. **B)** Análisis de la expresión por qPCR de las uniones de exones de *Tcfl5* e3/e4 (negro), e5/e6 (gris) y e4/e4l (blanco) en las líneas embrionaria de ratón *Tcfl5*^{-/-} e3-1, e3-2, e4l, e5 y e6. La línea roja representa la referencia de expresión de la línea Wt. **C)** Análisis de la expresión por qPCR de *Sox2* (gris) y *Nanog* (blanco) en las líneas embrionaria de ratón *Tcfl5*^{-/-} e3-1, e3-2, e4l, e5 y e6. La línea roja representa la referencia de expresión de la línea Wt. **D)** Arriba, análisis de la expresión por Western-blot de las proteínas NANOG, SOX2 y HSP90 en las líneas embrionaria de ratón *Tcfl5*^{-/-} e3-1, e3-2, e5 y e6. Abajo, cuantificación por densitometría del WB normalizando con la expresión de HSP90. En las gráficas se representa la media ± SEM de tres experimentos independientes y su estadística a la línea Wt (**p<0,01; ***p<0,001).

NANOG (Figura 32D). La expresión de NANOG se vio inducida en todas las líneas celulares. En cambio, la expresión de SOX2 se vio reducida en estas las mismas líneas.

4.3. La expresión de *Tcf15* se induce durante la diferenciación de células primordiales germinales (PGCs).

Los resultados obtenidos indican que la expresión de TCFL5 altera la expresión de las proteínas de pluripotencia, aunque sus efectos no sean visibles a nivel de ARNm. Esto podría repercutir en la capacidad pluripotente de estas mESC y por tanto alterar la capacidad de diferenciación hacia los distintos linajes. A partir de esta hipótesis, decidimos diferenciar las líneas embrionarias de ratón generadas hacia células primordiales germinales (PGCs) ya que, como hemos apuntado anteriormente, la mayor expresión de *Tcf15* se ha visto en testículo y se ha descrito su papel durante la espermatogénesis. Tras la diferenciación, se analizó por qPCR la expresión de las uniones de exones de *Tcf15*, marcadores de pluripotencia (*Sox2*, *Klf4*, *Nanog* y *Pou5f1*) y marcadores de diferenciación a línea germinal (*Dazl*, *Sohlh2* y *Stella*). Durante la diferenciación la expresión de las uniones de exones de *Tcf15* e3/e4, e4b y e5/e6 aumentó progresivamente hasta día 8 volviendo a los niveles basales a día final del experimento (Figura 33A). Con respecto a los marcadores de pluripotencia (Figura 33B), la expresión de *Sox2* disminuyó a día 4 para aumentar su expresión a día 8 y volver a los niveles basales a día 10. La expresión de *Klf4* se vio reducida alcanzando los niveles más bajos a día 4 y volviendo progresivamente a los valores basales a día 10. La expresión de *Nanog* disminuyó y se mantuvo baja

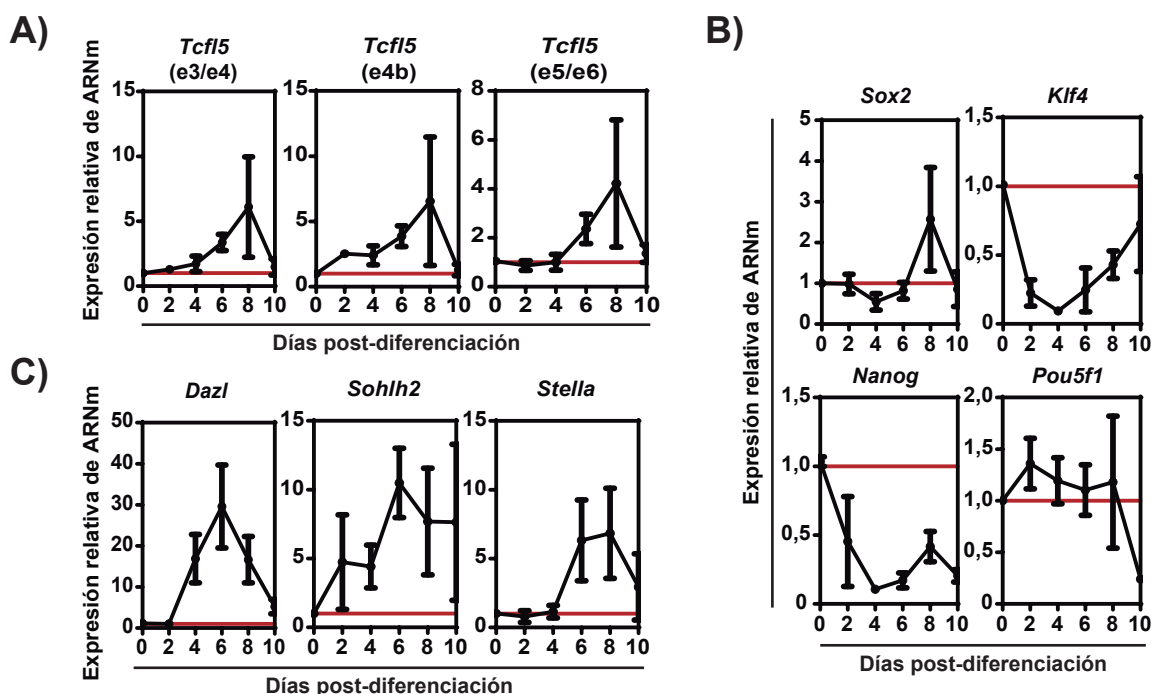


Figura 33. La expresión de *Tcf15* se induce en la diferenciación de las células embrionarias de ratón hacia progenitores de la línea germinal. Análisis de la expresión por qPCR de las isoformas de *Tcf15*, marcadores de pluripotencia y marcadores de diferenciación germinal durante el proceso de diferenciación in vitro hacia progenitores primordiales de la línea germinal de células embrionaria de ratón. **A)** Expresión de las uniones de exones e3/e4, e4/e4l y e5/e6 de *Tcf15*. **B)** Expresión de los marcadores de pluripotencia *Sox2*, *Klf4*, *Nanog* y *Pou5f1*. **C)** Expresión de los marcadores de línea germinal *Dazl*, *Sohlh2* y *Stella*. En las gráficas se representa la media \pm SEM de tres experimentos independientes.

y constante a partir del día 4. Por último, la expresión de *Pou5f1* se mantuvo inducida hasta día 6 a partir del cual disminuyó a sus niveles más bajos. Finalmente, los marcadores de diferenciación *Dazl*, *Sohlh2* y *Stella* aumentaron su expresión a lo largo del experimento alcanzando sus niveles máximos a día 6 (Figura 33C). Estos resultados indican que durante la diferenciación disminuye la expresión de los distintos genes de pluripotencia silenciando su expresión mientras que los genes de diferenciación a línea germinal incrementan durante el proceso. En el caso de la expresión de *Tcfl5*, este sigue el patrón de los genes de diferenciación sugiriendo un posible papel en este proceso pudiendo actuar como represor transcripcional de los genes de pluripotencia.

4.4. La deficiencia de TCFL5 inhibe la diferenciación a PGCs.

Por último, realizamos en las mESC *Tcfl5*^{-/-} el mismo proceso de diferenciación a PGCs para comprobar el efecto que la deficiencia de las isoformas de *Tcfl5* tiene sobre este proceso. Primero, analizamos el efecto de la delección de todas las isoformas utilizando la línea e3-2. No se realizaron los experimentos con la línea e3-1 debido a que en esta no silenció la isoforma *Tcfl5_e4l*. La expresión de las uniones de exones e3/e4 y e5/e6 fueron mucho menores durante todo el proceso de diferenciación, como era de esperar. (Figura 34A). Únicamente la unión de exones e4b presentó una mayor expresión que la línea Wt a día 10. Los marcadores de pluripotencia *Sox2*, *Klf4*, *Nanog* y *Oct4* (*Pou5f1*) presentaron patrones de expresión distintos en la línea e3-2 respecto a la línea Wt (Figura 34B). En general los patrones de expresión en la línea e3-2 fueron ligeramente inferiores al inicio de la diferenciación (día 2) al de la línea WT, similares en tiempos intermedios, pero aumentan mucho en los estadios finales (día 10), sugiriendo de nuevo un papel represor de *Tcfl5* en la expresión de estos genes. Con respecto a los marcadores de diferenciación *Dazl* y *Sohlh2* su expresión aumentó durante la diferenciación, aunque se mantuvieron más bajos que en el Wt en todo el proceso (Figura 34C). Únicamente la expresión de *Stella* fue la misma para ambas líneas celulares. Estos resultados parecen indicar que la deficiencia total de *Tcfl5* altera la diferenciación a línea germinal manteniendo los marcadores de pluripotencia elevados mientras que los marcadores de diferenciación disminuyen.

Una vez definido el papel de la deficiencia de todas las isoformas, continuamos el estudio con las líneas e5 y e6 en la diferenciación a línea germinal. La línea e5 expresa únicamente las isoformas *Tcfl5_e6Δ5* y *Tcfl5_e4l*. Por su parte, la línea e6 expresa las isoformas *Tcfl5_e4l* y *Tcfl5_e5*. Analizamos la expresión de la unión de exones e3/e4, e4b y e5/e6 (Figura 35A). La unión de exones e3/e4 que reconoce la expresión de todas las isoformas se mantuvo baja durante el proceso en la línea e6. Sin embargo, el patrón de expresión en la línea e5, aunque a niveles más bajos, fue igual que en la línea Wt indicando expresión de alguna isoforma. La unión e4b, específica para la isoforma *Tcfl5_e4l*, no se debería de ver afectada. Sin embargo, su expresión fue baja durante toda la diferenciación sugiriendo una posible regulación de su expresión por parte de alguna otra isoforma. La expresión de la unión e5/e6, específica para las isoformas delecionadas en estas líneas, se mantuvo en niveles bajos. En cuanto a los marcadores de pluripotencia, su expresión se mantuvo con un

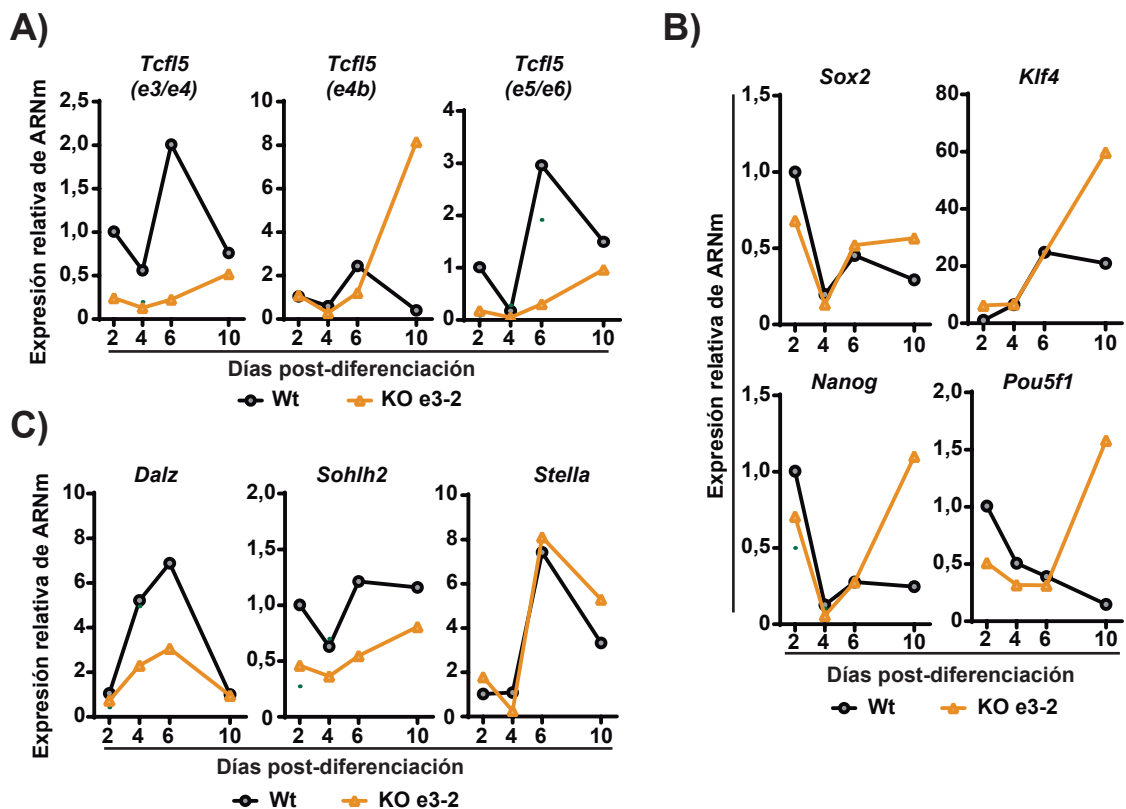


Figura 34. La deficiencia completa del gen *Tcf15* altera la pluripotencia y produce una inhibición de la diferenciación de las células embrionarias de ratón hacia progenitores de la línea germinal. Análisis de la expresión por qPCR de las isoformas de *Tcf15*, marcadores de pluripotencia y marcadores de diferenciación germinal durante el proceso de diferenciación in vitro hacia progenitores primordiales de la línea germinal de células embrionaria de ratón Wt (círculo negro) y *Tcf15*^{-/-} e3-2 (triángulo amarillo). **A)** Expresión de las uniones de exones e3/e4, e4/e4l y e5/e6 de *Tcf15*. **B)** Expresión de los marcadores de pluripotencia *Sox2*, *Klf4*, *Nanog* y *Pou5f1*. **C)** Expresión de los marcadores de línea germinal *Dalz*, *Sohlh2* y *Stella*.

patrón de expresión similar en general a la línea Wt (Figura 35B). *Sox2*, *Nanog* y *Oct4* (*Pou5f1*) no presentaron cambios significativos entre líneas con respecto al Wt aunque ambas líneas presentaron menores niveles de expresión desde el inicio. Por su parte, los niveles de *Klf4* si se vieron afectados de manera diferencial entre líneas. Las líneas e5 y e6 presentaron un adelanto en el pico de expresión con respecto a la línea Wt y disminuyeron sus valores a día final del experimento. En cuanto a los marcadores de diferenciación *Dalz*, *Sohlh2* y *Stella* (Figura 35C), todos ellos presentaron el mismo patrón de expresión en las líneas e5 y e6, pero la inducción producida en este proceso fue mucho menor que en la línea Wt. En este caso, los datos reflejan que la deficiencia o presencia específica de ciertas isoformas de *Tcf15* producen una menor expresión de los marcadores de diferenciación, aunque no producen cambios demasiado pronunciados en la expresión de los marcadores de pluripotencia.

Finalmente, terminamos el análisis del efecto de *Tcf15* sobre la diferenciación a línea germinal en la línea e4b. Esta línea presenta únicamente deficiencia en la expresión de la isoforma *Tcf15_e4l* expresando por tanto el resto de isoformas. Sin embargo, la expresión de las isoformas durante el proceso de diferenciación en esta línea celular siguió un patrón distinto al de la línea Wt (Figura 36A). El patrón de expresión para las uniones de exones e3/e4, e4b y e5/e6 se mantuvo bajo y

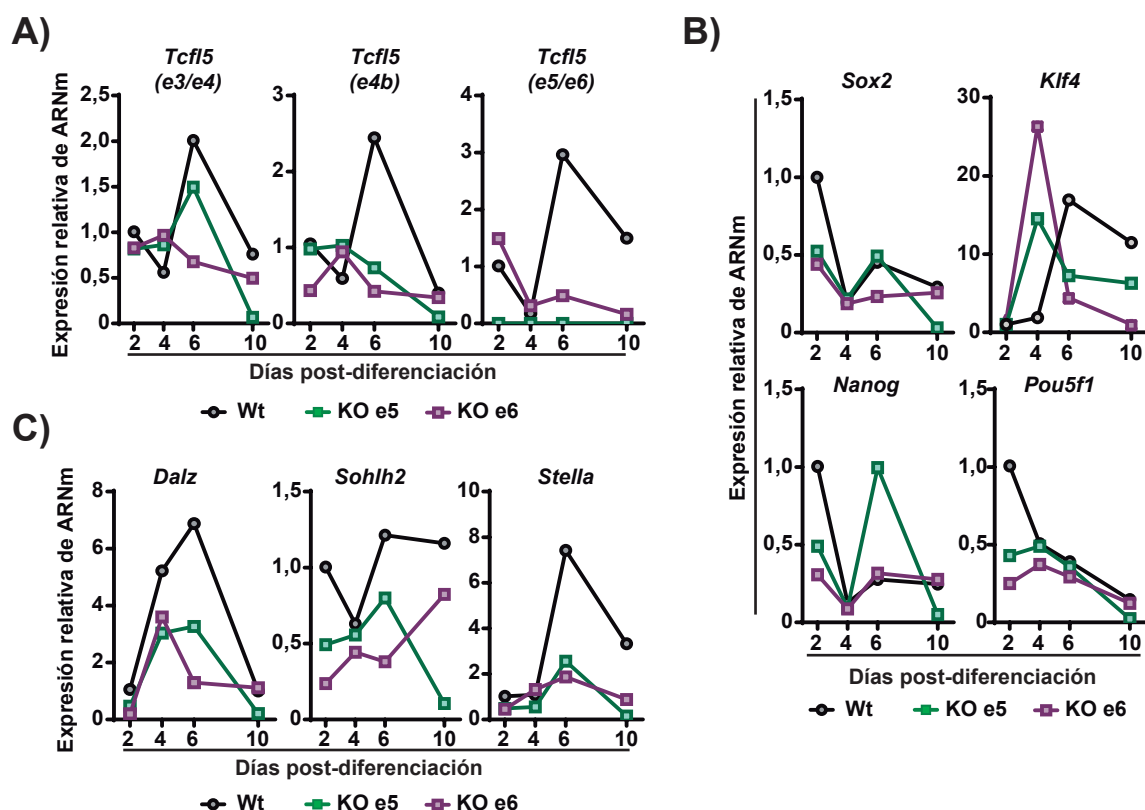


Figura 35. Las líneas *Tcf15*^{-/-} e5 y e6 no alteran la pluripotencia y retrasan la diferenciación de las células embrionarias de ratón hacia progenitores de la línea germinal. Análisis de la expresión por qPCR de las isoformas de *Tcf15*, marcadores de pluripotencia y marcadores de diferenciación germinal durante el proceso de diferenciación in vitro hacia progenitores primordiales de la línea germinal de células embrionaria de ratón Wt (círculo negro) y *Tcf15*^{-/-} e5 (cuadrado verde) y e6 (cuadrado morado). **A)** Expresión de las uniones de exones e3/e4, e4/e6 y e5/e6 de *Tcf15*. **B)** Expresión de los marcadores de pluripotencia *Sox2*, *Klf4*, *Nanog* y *Pou5f1*. **C)** Expresión de los marcadores de línea germinal *Dazl*, *Sohlh2* y *Stella*.

prácticamente constante a días 2, 4 y 6 aumentando su expresión a día 10 ligeramente por encima de la línea Wt, no presentando el pico de inducción de *Tcf15* característico a 6-8 días, sugiriendo un retraso en la diferenciación. En cuanto a los marcadores de pluripotencia *Sox2*, *Klf4*, *Nanog* y *Oct4* (*Pou5f1*) (Figura 36B), todos, excepto *Klf4*, presentaron el mismo patrón de expresión y completamente opuesto a la Wt. En la línea e4b, la expresión de estos genes fue baja en los primeros días y aumentó a día final. En el caso de *Klf4* no se observaron diferencias de expresión durante el proceso excepto al día final. Estos datos fueron similares a los obtenidos con la deficiencia de completa (e3-2). Con respecto a los marcadores de diferenciación *Dazl*, *Sohlh2* y *Stella* (Figura 36C), su expresión también presentó variaciones a lo largo del proceso. *Dazl* prácticamente no aumentó su expresión durante la diferenciación. *Sohlh2* tuvo un pico máximo de expresión a día 10 indicando un retraso en su inducción y *Stella* se indujo de nuevo durante la diferenciación, pero a niveles más bajo que la línea Wt. Por lo tanto, la delección específica de la isoformas *Tcf15_e4l* regula la expresión del resto de isoformas de *Tcf15* del mismo modo que altera la expresión de los marcadores de pluripotencia durante este proceso y produce una menor expresión de los marcadores de diferenciación.

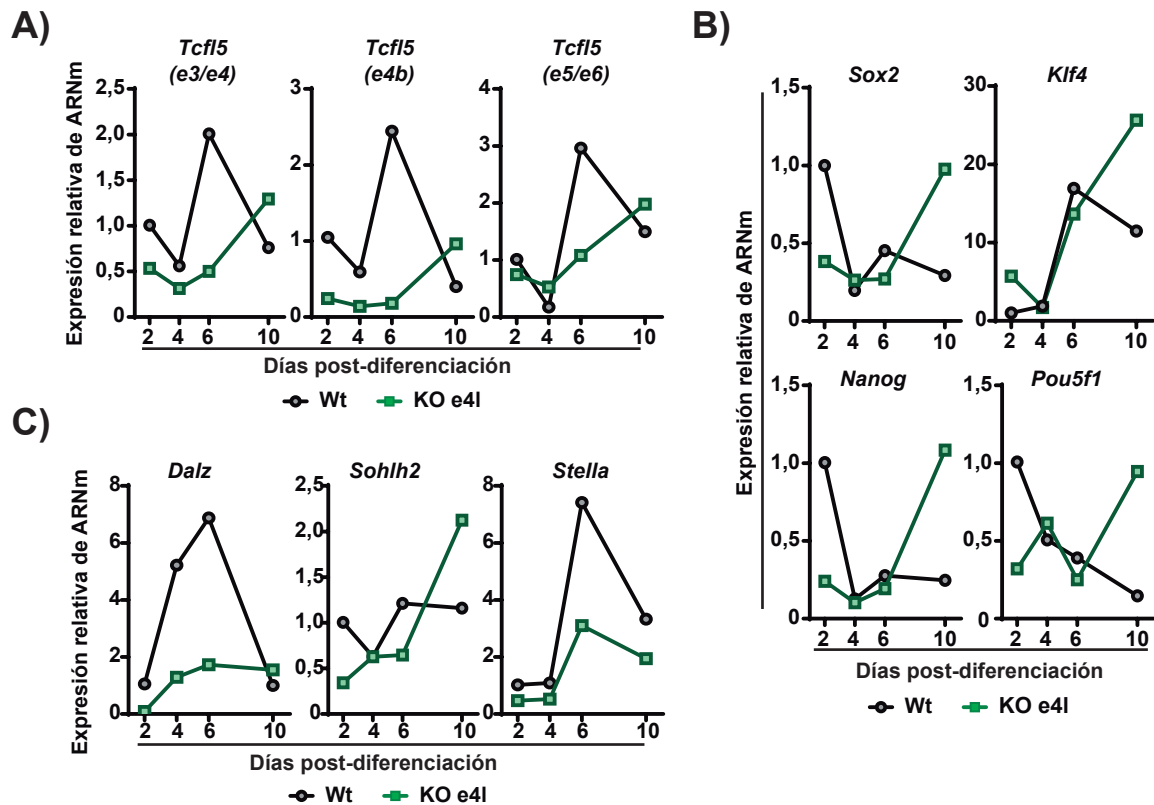


Figura 36. La deficiencia de la isoforma *Tcf15* *e4l* altera la pluripotencia y produce una inhibición de la diferenciación de las células embrionarias de ratón hacia progenitores de la línea germinal. Análisis de la expresión por qPCR de las isoformas de *Tcf15*, marcadores de pluripotencia y marcadores de diferenciación germinal durante el proceso de diferenciación in vitro hacia progenitores primordiales de la línea germinal de células embrionaria de ratón Wt (círculo negro) y *Tcf15*^{-/-} *e4l* (cuadrado verde). **A)** Expresión de las uniones de exones e3/e4, e4/e4l y e5/e6 de *Tcf15*. **B)** Expresión de los marcadores de pluripotencia *Sox2*, *Klf4*, *Nanog* y *Pou5f1*. **C)** Expresión de los marcadores de línea germinal *Dazl*, *Sohlh2* y *Stella*.

Discusión

Discusión

1. Expresión de *TCFL5* en tejidos humanos y de ratón.

Diversos autores han encontrado la expresión del gen *TCFL5* como específica de testículo [98] [99] [100]. No obstante, varios trabajos han descrito su función en otros tipos celulares y procesos como cáncer y sistema inmune [104] [111] demostrando que su expresión no es exclusiva de testículo. Nuestros datos muestran que, en humanos, el gen de *TCFL5* se expresa a nivel de mensajero en todos los tejidos y primeros estadios del desarrollo embrionario, encontrándose su mayor expresión en testículo seguido de cerebro.

En la literatura se han descrito varias isoformas del gen *TCFL5*. Gironés y colaboradores describieron la proteína CHA como una isoforma de *TCFL5* en la línea celular Jurkat, modelo para el estudio de leucemias de células T, pero expresada en prácticamente todos los tejidos humanos a nivel de proteína y ARNm [101] [103]. En 2013, Shi y colaboradores plantearon la posible existencia de varias isoformas para *TCFL5* en espermatogénesis [100]. Según su trabajo, la proteína que ellos caracterizaron difiere de la descrita por Siep y colaboradores [99] en secuencia, expresión y localización celular. Por otro lado, los datos de ciertos microarrays presentes en bases de datos muestran el uso de dos sondas para la detección de este gen. Sorprendentemente, encontramos que una de estas sondas no reconocía ninguna de las isoformas previamente descritas en la literatura, lo que plantea la cuestión de cómo y por qué se incluyó en esos arrays. Además, las nuevas técnicas de secuenciación han permitido anotar en diferentes bases de datos nuevos transcritos para este gen. Con esta información, hemos descrito en humanos al menos 6 isoformas para *TCFL5* (*TCFL5_E8*, *TCFL5_E8b*, *CHA*, *TCFL5_E6*, *TCFL5_E6b* y *TCFL5_E7*). Estas isoformas difieren en el primer o último exón, y 2 de ellas, las denominadas b, solamente en un aminoácido de diferencia, pero todas ellas presentan la región bHLH intacta por lo que consideramos que son potencialmente funcionales. De esta forma, hay 3 exones finales alternativos (E8, E6 y E7) y dos exones iniciales (E1 y E2b). Hemos confirmado que estas isoformas se encuentran expresadas en todos los tejidos a nivel de mensajero sin encontrar ninguna isoforma específica de tejido. Sin embargo, no podemos descartar algún tipo de control post-transcripcional que regule la expresión de la proteína en tejidos. No se ha descrito ningún dominio funcional en estos exones que diferencie posibles funciones entre ellas. No obstante, la distinta secuencia proteica puede permitir una diferente conformación permitiendo interactuar con diferentes proteínas, tener diferentes modificaciones post-traduccionales o incluso diferentes localizaciones celulares pudiendo llegar a proporcionar distintas propiedades a las isoformas [150].

Hemos planteado la posibilidad de que presenten dos promotores alternativos. Aunque esto precisa de un estudio más detallado, el análisis del promotor sugiere esta posibilidad localizándose la región promotora de *TCFL5* en los exones E1 y E2b coincidiendo con los inicios de transcripción

de estos dos exones. Resultados del laboratorio sin publicar (Ángel Gutiérrez-Nogués) confirmarían la existencia de estas 2 regiones promotoras. Además, presenta un promotor pausado. Este promotor pausado o bivalente se caracteriza por presentar modificaciones epigenéticas de las histonas activadoras y represoras al mismo tiempo y fue descrito en sus inicios en los promotores de genes de desarrollo embrionario y asociado a pluripotencia [151] [152]. Estudios recientes han demostrado que estos promotores se mantienen en los genes críticos para el desarrollo de la línea germinal masculina y femenina [153] [154]. Esto coincide con la expresión de *TCFL5* durante la espermatogénesis y desarrollo.

En ratón, la expresión del gen y composición de los transcritos es bastante distinta. La expresión del gen *Tcfl5* en tejidos adultos es específica de testículo y se encuentra también expresado en el periodo de preimplantación del proceso de desarrollo y en células embrionarias. Además, hemos descrito 5 isoformas para *Tcfl5* (*Tcfl5_e6*, *Tcfl5_e6b*, *Tcfl5_e6Δ5*, *Tcfl5_e4l* y *Tcfl5_e5*) todas ellas expresadas en testículo y células embrionarias. Estas isoformas difieren mucho de las isoformas de humano. Las isoformas *Tcfl5_e6Δ5* y *Tcfl5_e4l* no presentan isoformas similares en humano y no tienen la región bHLH completa por lo que sugerimos que su función como factor de transcripción se puede limitar a actuar como dominante negativo del resto de isoformas. La isoforma *Tcfl5_e5* utiliza el exón e5 como exón final y tiene un bHLH funcional. Esta isoforma sería similar a la isoforma *TCFL5_E6* de humano puesto que el exón E6 humano codifica para solo 2 aminoácidos siendo el resto de la secuencia un 3' UTR pudiendo ser su función similar en humanos y ratón. Por último, las isoformas *Tcfl5_e6* y *Tcfl5_e6b* son similares a las isoformas *TCFL5_E8* y *TCFL5_E8b*.

No hemos encontrado ninguna evidencia que soporte la presencia de algún transcrito que codifique para *Cha* en ratón. En humanos y en extractos de corazón y macrófagos de ratones se ha reportado la presencia de la proteína short-CHA (sCHA) importante y necesaria para el desarrollo de autoinmunidad cruzada en la infección por el parásito *T. cruzi* tanto en humanos como en ratón [101]. Este producto proteico en humanos comienza en una metionina del exón E4 humano por lo que podría comenzar su transcripción desde ese lugar, aunque en nuestro análisis no hemos encontrado ningún transcrito para esa forma. Sin embargo, la secuencia de ratón no presenta una metionina en esa posición. La proteína sCHAY CHA se producen en ratón por lo que se sugiere que sean un producto de proteólisis de la proteína *TCFL5*.

En ratón no existe el exón E7 y no se genera ninguna isoforma parecida. El exón E7 se encuentra en la zona del pseudogen *DPH3P1* de manera antisentido. Este pseudogen aparece sólo en primates y es probable que durante la evolución se haya incorporado por *splicing* alternativo a la secuencia de *TCFL5* y utilizado como exón final alternativo [155]. La aparición de este exón coincide evolutivamente con el desarrollo de los procesos cognitivos. Además, la alta expresión de *TCFL5* en cerebro en humano, pero no en ratón, podría indicar un papel importante de este gen en procesos cerebrales humanos.

Por último, la expresión en testículo y en desarrollo embrionario tanto en humano como en ratón indica un papel principal y evolutivamente conservado en estos tejidos. Sin embargo, la diferencia de expresión en tejidos adultos de humanos y ratón, así como las diferentes isoformas encontradas en ambos, principalmente *CHA* y *TCFL5_E7* sugiere que, durante la evolución, se ha producido cambios en la expresión añadiendo nuevas funciones a este gen indicando una evolución muy reciente del mismo.

2. Expresión de *TCFL5* en cáncer de colon.

Algunos trabajos sugerían un papel importante de *TCFL5* durante el desarrollo de cáncer colorrectal. Por ello, una parte de nuestro trabajo se centró en este tipo de cáncer. La expresión de *TCFL5* es mayor en carcinomas y se relaciona con la inestabilidad cromosómica de la región 20q [116]. Esta alteración cromosómica se ha encontrado en un alto porcentaje de carcinomas, aunque también está presente en menor medida en adenocarcinomas. Los genes que se encuentran en esta región se han asociado principalmente a una progresión del estado de carcinoma sin descartar un posible papel en los primeros estadios del inicio del tumor [156]. Nuestros análisis de las bases de datos muestran que efectivamente hay una sobreexpresión de las isoformas del gen *TCFL5* en carcinoma de colon con respecto a un colon sano y que esta expresión se relaciona con un aumento del número de copias de este gen. Estos resultados indican que la inducción del gen es consecuencia de esta inestabilidad cromosómica y no de un aumento de la actividad de sus vías de señalización. Por tanto, la expresión aberrante y continua de *TCFL5* puede jugar un papel importante en la progresión a carcinoma, incluso en el inicio del tumor.

En la colección de tumores estudiados un bajo porcentaje de casos presentaron mutaciones. Ninguna de ellas se encontraba en la región funcional bHLH y todas ellas fueron mutaciones sin sentido. A día de hoy, no se ha descrito ninguna mutación de este gen asociado a cáncer de colon ni se ha estudiado su papel en la actividad o estabilidad de la proteína.

3. Auto-regulación de *TCFL5* y *CHA*.

Durante el desarrollo experimental observamos que podía existir una regulación de la expresión de estas isoformas por ellas mismas. La autorregulación de isoformas se ha estudiado por ejemplo en el caso del factor de transcripción PAX6. PAX6 presenta dos isoformas con funciones regulatorias distintas y se ha observado que la transfección de una de ellas incrementa la expresión de la otra isoforma y viceversa [170]. Nuestros datos mostraron que *TCFL5* indujo la expresión de *CHA* y viceversa indicando una regulación recíproca. Esta regulación puede observarse en la formación de MCTS. La inducción del gen tras la formación de MCTS de las líneas que sobreexpresan estas

isoformas fue mucho mayor sugiriendo un efecto sinérgico de la sobreexpresión y la formación de MCTS. Además, ambas isoformas se unían físicamente a sus promotores por lo que esta regulación tiene lugar de manera directa. De este modo, al tener funciones diferentes en la célula, esta autorregulación serviría como vía para mantener un ratio de expresión TCFL5/CHA constante por lo que un desajuste de este promueve cambios en el fenotipo. El mantenimiento del ratio de expresión de isoformas es un proceso importante y regulado. Por ejemplo, en el caso de las isoformas LAP y LIP del gen C/EBP β , se ha descrito que la desregulación del ratio determina la progresión del ciclo celular [171]. Varios de nuestros datos podrían explicarse por el ratio TCFL5/CHA en cada una de las ocasiones experimentales. Por eso los resultados al sobreexpresar TCFL5 o CHA en líneas como HCT116 que expresan bastante CHA parecen contradictorios a los de la línea HT29 que apenas expresan CHA (Figura 37).

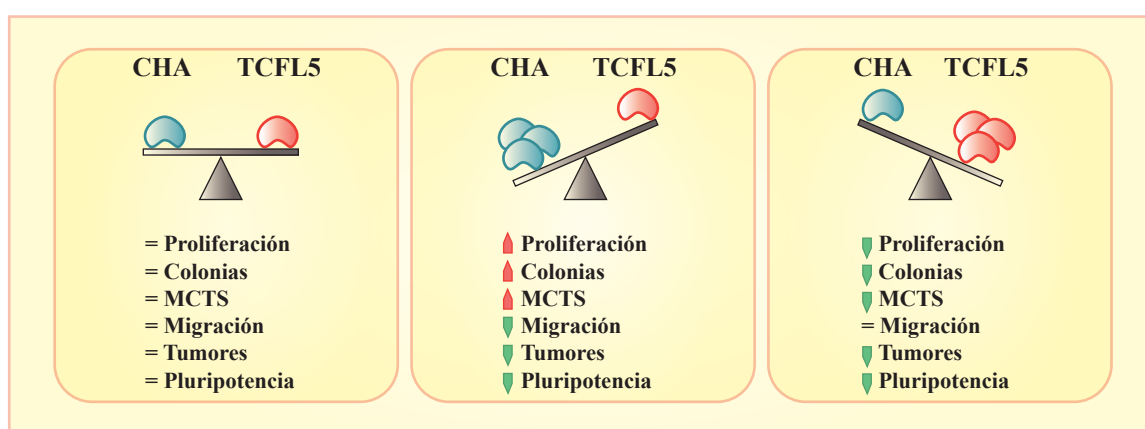


Figura 37. Esquema resumen de los efectos de la sobreexpresión de TCFL5 y CHA sobre el fenotipo de células de cáncer de colon.

4. Papel diferencial en el fenotipo celular producido por las isoformas TCFL5 y CHA.

Para el estudio del papel de las isoformas en cáncer de colon utilizamos las líneas celulares HT29 y HCT116 debido a su diferencia de expresión basal de TCFL5. Al describir las isoformas presentes en estas líneas, observamos que la línea celular HCT116 expresa las isoformas *TCFL5* y *CHA* mientras que la línea HT29 tiene una expresión mayoritaria de *TCFL5*, pero no de *CHA*. Estas líneas celulares provienen de distintos estadios de la progresión colorrectal siendo la línea HCT116 de un estadio más avanzado (Dukes D) que la línea HT29 (Dukes C) [157]. Esta diferencia de expresión de las isoformas en estas dos líneas celulares sugiere que podría tener lugar un cambio del patrón de expresión de las isoformas durante el desarrollo del cáncer colorrectal y que la expresión de CHA correlacionaría con un fenotipo tumoral más avanzado o metastásico.

En la línea HT29, la sobreexpresión de TCFL5 no produjo ningún efecto en el fenotipo de la célula mientras que la sobreexpresión de CHA si lo tuvo. La ausencia de efecto de TCFL5 probablemente se deba a que esta línea celular presentó una alta expresión basal de esta isoforma por lo que un aporte exógeno de ella no produce ningún efecto. Sin embargo, la sobreexpresión de CHA si produjo efecto debido a su baja expresión basal en esta línea celular, que alteraría la expresión de

posibles genes diana de CHA. La sobreexpresión de CHA aumentó la proliferación y la capacidad clonogénica *in vitro*, mientras que redujo la tumorigénesis *in vivo*. En la literatura se ha descrito en esta misma línea celular que el silenciamiento de TCFL5 (que no sabemos si también puede afectar a CHA) redujo la viabilidad celular [115] la proliferación y la formación de MCTS [111]. Estos datos coinciden con los observados con la sobreexpresión de CHA. Sin embargo, en nuestro caso, los efectos de estos silenciamientos fueron distintos entre sí y solamente uno de ellos tuvo los efectos previamente descritos. Este silenciamiento se llevó a cabo con shARNs que mapeaban en diferentes regiones del gen y por tanto silencian diferentes isoformas. De esta forma, el shARN específico para las isoformas TCFL5 y CHA fue el que redujo la proliferación mientras que el shARN para todas las isoformas no tuvo efecto. Esto sugiere un papel del resto de isoformas, posiblemente opuesto, en esa línea de cáncer de colon. En conjunto estos datos indican que hay que ser más cuidadosos al implicar a un gen en un fenotipo mediante este tipo de análisis cuando hay diversidad de isoformas.

En la línea HCT116 se produjo el fenotipo opuesto. La sobreexpresión de TCFL5 redujo la proliferación *in vitro*, capacidad clonogénica y crecimiento tumoral *in vivo* mientras que CHA no produjo cambios en estos procesos, pero sí disminuyó la migración *in vitro*. En esta línea el silenciamiento con shARNs fue poco efectivo por razones desconocidas, por lo que se generó la delección completa de todo el locus de TCFL5 por CRISPR-Cas. Esta produjo una inhibición de la proliferación *in vitro*, una reducción del tamaño tumoral *in vivo* y un aumento de la migración *in vitro*. Los resultados de la sobreexpresión de una isoforma y delección resultan en primera instancia similares. No obstante, los efectos de la delección coinciden con los opuestos de CHA observados en la línea celular HT29. La línea celular HCT116 presentó una expresión basal de ambas isoformas, aunque mayor para CHA que para TCFL5. Por esta razón, se observa un mayor efecto de la sobreexpresión de la isoforma TCFL5, mientras que al silenciar predominaría el efecto de la falta de CHA.

Los análisis de ciclo celular realizados en líneas HCT116 muestran que los efectos observados por las líneas sobreexpresadas y silenciadas se producen en puntos distintos del ciclo celular. p21 (WAF1/Cip1) y p27 (KIP1) son proteínas inhibidoras de ciclinas dependientes de quinasas que se asocian con una parada del ciclo celular en fase G0/G1 [136]. La expresión de ambas se ve alterada por la expresión de TCFL5 o CHA y por la delección completa. Aunque los datos son preliminares, en HCT116 la sobreexpresión de TCFL5, que retarda el crecimiento, indujo una bajada más rápida de p27, un retraso en la subida de p21 y acumulación de células en fase G2/M. Mientras que CHA, que tiene menos efectos en crecimiento, tiene en general siempre menos niveles de ambas. Estos resultados indican que la sobreexpresión de TCFL5 puede afectar al punto de control G2/M o impedir la mitosis. En las líneas HCT116 TCFL5^{-/-}, que crecen menos *in vitro*, se observó un aumento del número de células en fase G0/G1 y una disminución de células en fase S sin alterarse el número de células en fase G2/M. En esta línea, la expresión de p27 disminuyó en el tiempo respecto a la línea control mientras que la expresión de p21 aumentó. No obstante, nuestros resultados sugieren que

la deficiencia completa de TCFL5 y CHA produce cierta parada del ciclo celular en fase G0/G1 o impide la replicación del ADN.

En este sentido, nuestros datos del interactoma muestran que TCFL5 se une a la proteína *Proliferation Cell Nuclear Antigen* (PCNA). PCNA es una proteína necesaria para la replicación [158] y su inhibición impide la replicación celular tanto *in vitro* como *in vivo* [159] [160]. TCFL5 podría interaccionar con PCNA y formar parte del complejo de replicación por lo que su ausencia retrasaría el progreso de la síntesis de ADN. En la misma línea, estudios en líneas de cáncer de colon Caco2 y HT29 demostraron que el silenciamiento de TCFL5 produjo una reducción de la viabilidad celular tras el bloqueo de la síntesis de ADN con el tratamiento con 5 fluorouracilo (5FU) sugiriendo un papel en este proceso [115]. Además, los datos del interactoma también mostraron que TCFL5 se une a la proteína DDB1. DDB1 es una proteína involucrada en la vía de reparación del daño en el ADN por escisión de nucleótidos (NER) que se une al daño producido por luz ultravioleta [161]. La unión de TCFL5 a DDB1 sugiere un papel en el control del daño en el ADN que puede afectar a la progresión del ciclo celular.

La formación de MCTS se estudia también para medir la capacidad tumoral de las células de tal forma que una célula capaz de formar MCTS se considera que presenta un fenotipo más agresivo [162]. El estudio publicado por Dardousis y colaboradores muestra una inducción de la expresión de *TCFL5* en los MCTS y que su silenciamiento reduce su formación [111]. Nuestros datos coinciden en parte ya que revelan que la inducción se observa en todas las isoformas del gen y que el locus *TCFL5* es absolutamente necesario para la formación de MCTS. Sin embargo, el papel individual de cada una de las isoformas parece ser distinto incluso opuesto. No obstante, la deficiencia completa de *TCFL5* y *CHA* impide la formación de MCTS.

TCFL5 inhibe la proliferación y MCTS mientras que *CHA* lo favorece promoviendo un fenotipo más tumorogénico. Por otra parte, *CHA* inhibe la migración celular y el crecimiento de tumores *in vivo*. Estos resultados parecen en primera instancia contradictorios. La mayor proliferación *in vitro* se asocia con un aumento del tamaño de los tumores [163] [164]. No obstante, la menor migración revela falta de movilidad y por tanto cambios en los procesos asociados a este fenómeno. Esto se relaciona con el proceso de transición epitelio-mesénquima (EMT). La transición EMT es un proceso en el cual las células epiteliales pierden sus uniones y polaridad produciendo células mesenquimales con capacidad migratoria y de invasión [165]. Este proceso es necesario para que una célula cancerosa migre de su lugar de origen a otro. Una vez asentadas allí, tiene lugar la transición mesénquima-epitelio (MET) para proliferar en el nuevo lugar [166]. Diversas proteínas se han asociado a este proceso. Por ejemplo, las proteínas de adhesión están implicadas y se encuentran frecuentemente alteradas en cáncer produciendo un aumento de la proliferación y la diseminación del tumor [167]. Sin embargo, para algunas de estas proteínas se ha descrito que su sobreexpresión disminuye la proliferación y formación de tumores y aumenta la migración [168] [169]. Por tanto,

la formación de tumores más pequeños debido a la sobreexpresión de CHA puede explicarse no por una menor proliferación, sino por una peor implantación de las células una vez han sido inoculadas y por tanto generar un retraso en el comienzo de la proliferación. Esto sugiere que la sobreexpresión de CHA podría favorecer el proceso MET y no el EMT generando células más proliferativas, pero con menos capacidad de migración y colonización. También es posible que la expresión de CHA resulte en un fenotipo más “*stem*” asociado a más tumorigenicidad, pero menor crecimiento.

Los resultados del crecimiento de tumores xenoonjertados *in vivo* no correlacionan directamente con los datos de proliferación *in vitro* de las diferentes líneas. En cualquier caso, la deficiencia completa de todas la isoformas reduce drásticamente el crecimiento de los tumores en xenotransplantes e impide su capacidad de formar MCTS. En conjunto nuestros resultados indican que TCFL5 y CHA tienen un papel importante en la progresión del cáncer de colon especialmente en procesos no estrictamente proliferativos y con funciones muy diferentes entre ambas isoformas.

5. Interacción de TCFL5 con NFKB2 y BCL2.

Se ha descrito que las isoformas generalmente suelen compartir menos del 50% de las interacciones y comportarse más como proteínas distintas que como isoformas menores [172]. Nuestros datos muestran que un alto porcentaje de las proteínas obtenidas como candidatas para interactuar con TCFL5 y CHA fueron distintas. Esto reveló que la red de interacción de ambas isoformas no es la misma por lo que sus funciones en la célula son potencialmente diferentes.

De las proteínas obtenidas, validamos a la subunidad p100 de NFKB2 como proteína de interacción de TCFL5, pero no de CHA. En cáncer, dependiendo del tipo de tumor y línea celular, la función de NFKB2 se ha descrito como anti-tumoral o pro-tumoral [173]. Estudios en cáncer de endometrio y linfoma de Burkitt han descrito que una mayor expresión de RelB y p52 produce una mayor proliferación, migración e invasión [139] [174] [175]. No obstante, otros estudios han demostrado que un aumento de RelB disminuye la proliferación y la formación de tumores aumentando los niveles de expresión de p53 y p21 [176]. Sin embargo, el papel de NFKB2 en cáncer de colon es desconocido. La actividad NFKB2 en la línea HCT116 es baja por lo que decidimos activar la vía utilizando el ligando del receptor Fn14 TWEAK, una proteína que también se ha implicado en cáncer [177]. Nuestros datos revelan que las isoformas de TCFL5 regulan los niveles globales de proteína NFKB2 y, sobretudo su deficiencia, disminuye los niveles basales de actividad NFKB2 inhibiendo completamente la actividad inducida por TWEAK. La reconstitución de las líneas HCT116 TCFL5^{-/-} con TCFL5 no restaura más que parcialmente la actividad de NFKB2 inducida por TWEAK, mientras que la reconstitución con CHA todavía resulta en mayor inhibición indicando que la actividad se vio afectada de manera positiva con TCFL5 y negativa con CHA. Pero también indica que alguna otra isoforma de TCFL5 debería estar implicada en la reconstitución completa de la actividad NFKB2 en HCT116 TCFL5^{-/-}. En cualquier caso, revelan la importancia de TCFL5 en la activación de NFKB en

líneas de carcinoma de colon en respuesta a TWEAK.

Se ha descrito que el efecto del tratamiento con TWEAK es dependiente de línea celular de carcinoma de colon [178]. En líneas de cáncer de colon promueve la apoptosis y mayores niveles de TWEAK se asocian a una mayor supervivencia de los pacientes y a una menor invasión celular [179] [141] [142] [180]. Además, la inhibición del receptor Fn14 con tratamientos con anticuerpos para el receptor disminuyen la proliferación y los tumores xenoinjertados [181]. Por otra parte, hay estudios que relacionan la disminución de la proliferación y la invasión a una activación directa de NF κ B2 por la activación por TWEAK [182]. Sugerimos que TCFL5 podría cooperar directamente en la función de NF κ B2 uniéndose a él y por tanto promoviendo una menor proliferación mientras que CHA parece tener el efecto contrario, aunque probablemente en este caso de modo indirecto.

Es posible que el complejo TCFL5/NF κ B2 regule la expresión de *BCL2*. Hemos visto que *BCL2* es una diana transcripcional de TCFL5 y que su sobreexpresión en células HCT116 resulta en una menor inducción de *BCL2*. Este es también diana transcripcional de NF κ B2 aumentando su expresión [139] [174]. Sin embargo, los mayores niveles de NF κ B2 no siempre se asocian a un aumento de *BCL2*, sino también a una disminución [183]. En cáncer de colon, la expresión de *BCL2* se correlaciona negativamente con la agresividad y recurrencia del tumor [184]. Sin embargo, también se ha descrito que la delección de *BCL2* favorece la apoptosis aumentando la supervivencia al cáncer [185]. Por ello, planteamos la posibilidad de que TCFL5 se una a NF κ B2 y juntos repriman la expresión de *BCL2* y, por tanto, favorezcan la apoptosis e inhiban la proliferación.

6. TCFL5 y CHA en la vía de WNT y NOTCH1.

Las mutaciones en APC se producen en los primeros estadios del cáncer de colon. Una pérdida de la actividad APC conlleva un aumento de la actividad β -catenina/TCF4 [186] por lo que se ha asociado directamente a la vía de Wnt en el desarrollo del cáncer colorrectal [187]. Tanto el silenciamiento de TCF4 como de β -catenina reducen la formación de tumores, aumenta la apoptosis y disminuye la sensibilidad a quimioterapéuticos [188]. Además, la inhibición del complejo β -catenin/TCF4 produce los mismos efectos junto con una disminución de la formación de tumores [189]. En modelos de ratón la delección de *Tcf4* se ha observado que reduce la formación de tumores [190]. Hemos observado que el factor de transcripción TCF4 se une al promotor de *TCFL5* y que un aumento de β -catenina y del factor de transcripción TCF4 producen un aumento de la expresión de *TCFL5*. Esto indica que la expresión de TCFL5 está controlada por la vía de Wnt por lo que puede tener un papel relevante en el desarrollo del cáncer colorrectal. Debido a las funciones diferentes de las isoformas TCFL5 y CHA en la tumorigénesis de cáncer de colon, será necesario discriminar entre las isoformas para determinar la regulación de estas por Wnt y relacionarla con el crecimiento tumoral.

Otra de las vías implicada en cáncer colorrectal es la vía de NOTCH. *TCFL5* se ha descrito como una diana de NOTCH1 en células Jurkat [104]. Nuestros estudios muestran que la inactivación de la vía de NOTCH produce una disminución de la expresión de *TCFL5* indicando que efectivamente esta vía regula su expresión. La expresión de NOTCH1 se relaciona con un peor pronóstico del cáncer [191] y su sobreexpresión en células de cáncer de colon produce un fenotipo más tumoral aumentando la proliferación *in vitro* e *in vivo* y reduciendo la apoptosis mientras que su silenciamiento tiene el efecto contrario [192]. Además, la expresión de NOTCH1 se encuentra inducida en células CSCs y previene la apoptosis de esas células reprimiendo p27 [193]. Al igual que la vía de Wnt, la activación de la vía de NOTCH se asocia a una mayor tumorigénesis. De nuevo, será necesario discriminar entre las isoformas para determinar cuál de ellas se induce por NOTCH. El control de la expresión de *TCFL5* por dos de las principales vías implicadas en el desarrollo del cáncer colorrectal demuestra la relevancia que el estudio de este factor de transcripción y sus isoformas pueden tener en este proceso.

7. *TCFL5* y *CHA* en CSCs y pluripotencia en líneas celulares de cáncer de colon.

Debido al efecto de las isoformas sobre la tumorigenicidad de las células, planteamos la posibilidad de que estos factores de transcripción estuvieran favoreciendo o impidiendo la generación de CSCs. Células que presentan una mayor actividad *ALDH1* presentan mayor tumorigenicidad [194] [195]. Igualmente, la expresión de *LGR5* se asocia a una mayor proliferación, invasión y formación de tumores [196] [197] y la expresión de *CD133* también se ha asociado a la presencia de CSCs y al aumento de la tumorigénesis [198] [147]. Por el contrario, *EpHB2* se encuentra poco expresado en células diferenciadas [199], y su pérdida de actividad favorece la progresión tumoral [200]. Nuestros datos muestran que no hay diferencias en cuanto a estos marcadores se refiere, indicando que *TCFL5* y *CHA* no afectan significativamente a la generación de CSCs.

Los marcadores de pluripotencia *SOX2*, *NANOG* y *KLF4* también se han relacionado con las CSCs [201]. Hemos observado que, tanto *TCFL5* como *CHA*, se unen a la región promotora de *SOX2* reprimiendo la actividad de esta. Además, la expresión de *SOX2* se encontró inducida en ausencia de ambas isoformas en HCT116 *TCFL5*^{-/-}. revirtiéndose este efecto tras la recuperación de la expresión de *TCFL5* y *CHA*. Sin embargo, en células WT, *TCFL5* indujo la expresión de *SOX2* y la actividad de su promotor. Esta diferencia de función de *TCFL5* sobre la expresión de *SOX2* en células WT o en células deficientes en todo el locus *TCFL5* sugiere una implicación de las otras isoformas de *TCFL5*, incluida *CHA*, en estos efectos. Es posible que algunos efectos de *TCFL5* dependan de la presencia (o inducción) o no de otras isoformas del locus ya que se regulan entre ellas y no es descartable la presencia de heterodímeros de las distintas isoformas para llevar a cabo una función. Se conoce que los factores bHLH son muy promiscuos y pueden formar homo y heterodímeros entre sí y con otras bHLH.

En cáncer, la deficiencia en *SOX2* se ha relacionado con una disminución de los tumores y

una menor proliferación y viabilidad celular [202]. En cáncer de colon, su expresión se ha asociado a un aumento de los marcadores de CSCs [203] [204] y a un peor pronóstico de los pacientes, promoviendo un aumento de la migración, invasión, proliferación y favoreciendo la progresión del cáncer [205] [206]. Sin embargo, aunque la sobreexpresión de SOX2 aumenta los marcadores de CSCs, también disminuye la migración y la proliferación [203]. Estudios en líneas HT29 y HCT116 han demostrado que, tanto la inhibición de SOX2 como su sobreexpresión, reducen la proliferación celular mientras que la viabilidad celular y la formación de tumores aumenta con su sobreexpresión tras el tratamiento de quimioterapéuticos [204]. Estos estudios demuestran que, aunque la expresión de SOX2 se asocia a un peor pronóstico, sus efectos sobre el fenotipo tumoral no están claros. La expresión de SOX2 en nuestras líneas se relaciona con un fenotipo menos agresivo. No obstante, se han observado distintas localizaciones celulares para SOX2 en cáncer de colon encontrándose en el citoplasma o en el núcleo [207] [208]. SOX2 es un factor de transcripción y por tanto su principal actividad la desarrolla en el núcleo. Nuestros estudios de actividad de SOX2 no fueron concluyentes debido a un mal funcionamiento del sistema. Son necesarios más estudios para dilucidar si SOX2 se encuentra activo transcripcionalmente o no en nuestras células y relacionarlo con el crecimiento celular y tumoral.

La expresión de NANOG en cáncer también se ha asociado al desarrollo tumoral [209]. Células con una mayor expresión de NANOG presentan un aumento de los marcadores de CSCs y su sobreexpresión produce un fenotipo más tumorogénico siendo esencial para el desarrollo tumoral [210] [211]. En líneas de cáncer de colon, cJUN y TCF4 promueven un fenotipo de CSCs a través del control de la expresión de NANOG [212]. Nuestros datos indican un control de la expresión por parte de TCFL5 y CHA siendo represores de la expresión y actividad de NANOG. Sin embargo, la expresión no coincide con el fenotipo menos tumorogénico esperado. La localización celular de este factor de transcripción en líneas de cáncer de colon y tejido tumoral es nuclear o citoplasmática dependiendo de la línea celular [213] [214]. Nuestros estudios de actividad transcripcional de NANOG revelan que únicamente la sobreexpresión de TCFL5 produce una menor actividad coincidiendo con los efectos fenotípicos, indicando que este se encuentra en el núcleo. Es probable que NANOG requiera estar en el núcleo para realizar su papel tumoral y por tanto ser más relevante la localización que la expresión.

El factor de transcripción KLF4 está involucrado en proliferación, diferenciación y apoptosis considerándose un gen supresor de tumores. Nuestros datos revelan que tanto TCFL5 como CHA son reguladores transcripcionales de KLF4 uniéndose a su promotor y regulando de manera negativa y positiva su transcripción, respectivamente. Sin embargo, la delección de todas las isoformas produce también una reducción en la expresión de KLF4 sugiriendo el efecto prioritario de CHA. En intestino, KLF4 se encuentra expresado en células no proliferativas donde se ha observado que el silenciamiento de este gen conlleva una activación de genes de la vía de Wnt y una reducción de genes reguladores de diferenciación [215]. Su expresión controla la parada del ciclo celular a

través del control de expresión de p21 [216] [217] [218]. En cáncer de colon, la expresión de KLF4 correlaciona negativamente con la progresión y la diferenciación celular [219] [220] [221]. Hemos observado que TCFL5 se induce en cáncer de colon y por tanto correlaciona negativamente con la expresión de KLF4. Sin embargo, esta expresión no explica el fenotipo observado en nuestras células. La sobreexpresión de este factor de transcripción en células de cáncer de colon conlleva a una reducción de la capacidad clonogénica, migración, invasión y tumorigénesis in vivo [222]. No obstante, KLF4 actúa como oncogén en condiciones de inactivación de p21 o RAS [223] y, en ciertas células de cáncer de colon, KLF4 induce la formación de MCTS, migración, invasión y formación de tumores aumentando la expresión de marcadores de CSCs [224]. Esto sugiere que el papel de KLF4 depende del contexto molecular de la línea celular. La baja expresión de KLF4 observada en nuestras células, si correlaciona con una menor proliferación y capacidad de formar tumores. Las líneas HT29 y HCT116 presentan distintas mutaciones en p53 y RAS así como reorganizaciones genómicas que podrían afectar a la funcionalidad de KLF4 [225] [157]. Además, estas líneas celulares presentan metilaciones y mutaciones en KLF4 que afectan a su control transcripcional de p21 y podrían alterar su función [219]. Recientemente han aparecido nuevas isoformas para este gen que se han asociado al desarrollo tumoral y con un efecto opuesto a la isoforma canónica [226], por lo que será necesario discriminar entre estas isoformas.

En conjunto NANOG y SOX2 parecen ser reguladas negativamente por el locus TCFL5 en HCT116 TCFL5^{-/-}. mientras que KLF4 seguiría un proceso distinto y que estos procesos dependerían de las isoformas de TCFL5, pero quizás no exclusivamente de CHA o TCFL5. Esto sugiere de nuevo que el balance de las isoformas es clave para el efecto final observado.

8. Expresión de TCFL5 durante el desarrollo embrionario temprano.

El control de SOX2, NANOG y KLF4 por parte de TCFL5 y CHA sugiere un papel de estos en pluripotencia. Además, en análisis de los estudios con embriones de ratón donde se demostró que el silenciamiento de *Oct4* en los estadios de 1 o 2 células impedía la progresión del desarrollo embrionario [123], observamos que dicho silenciamiento inhibía la expresión de *Tcfl5* indicando un control transcripcional por parte de OCT4 durante la primera ola de activación del cigoto. Hemos observado que, tanto en humano como en ratón, la expresión de TCFL5 se induce durante el desarrollo embrionario sugiriendo un papel importante en este proceso. Además, en humanos, los microarrays muestran que la isoforma *TCFL5_E6* presenta el patrón opuesto indicando que probablemente su expresión sea de procedencia materna mientras que la isoformas *TCFL5_E8* o *CHA* se inducen durante la primera ola de activación del cigoto. En ratón, hemos observado que las diferentes tinciones en inmunofluorescencia para distintas regiones de TCFL5 muestran un patrón de expresión distinto principalmente en estadio de 2 células. En este estadio, el exón e1 se encuentra excluido del núcleo mientras que los exones e4 y e5 son nucleares mayoritariamente. Estos datos sugieren que, de alguna manera, el exón e1 es procesado en este estadio quedándose fuera del núcleo. Aunque TCFL5

es un factor de transcripción y por tanto su expresión se considera nuclear, en espermatogénesis la presencia de este factor de transcripción se ha encontrado fuera del núcleo colocalizando con los centriolos [100]. Por otro lado, la expresión de TCFL5 en estadio de dos células sugiere también la presencia de un producto materno observándose su expresión en este estadio. Análisis de RNA-Seq o experimentos de *single molecule-FISH* (smFISH) serán necesarios para dilucidar la expresión de las distintas isoformas durante este proceso en ratón o si existe un procesamiento proteolítico que excluya al exón e1 del núcleo.

Por otro lado, hemos descrito la presencia de todas las isoformas en células embrionarias de ratón y hemos observado que, la deficiencia de la isoforma *Tcfl5_e4l* provocó una disminución del resto de isoformas sugiriendo un papel controlando su expresión. Además, en contra de lo que ocurre en humanos en cáncer de colon, la deficiencia de TCFL5 produjo una disminución de SOX2, pero un aumento similar en la expresión de NANOG. Estas diferencias observadas en el control de la expresión de SOX2 en humano y en ratón probablemente se deban a cambios en la red transcripcional de estas células. Sin embargo, durante la diferenciación de mESCs se observó que la expresión *Sox2*, *Nanog*, *Oct4* y *Klf4* presentaban un patrón inverso al de *Tcfl5*, sugiriendo una regulación negativa.

9. TCFL5 en diferenciación a línea germinal.

Algunos estudios ya han mostrado a TCFL5 asociado a procesos de diferenciación o reprogramación. Ensayos de reprogramación celular de fibroblastos embrionarios de ratón (MEFs) a células neuronales inducibles (iN) mostraron una inducción de *Tcfl5* durante este proceso. No obstante, la expresión por si sola de TCFL5 en MEFs no produjo la reprogramación, a diferencia de la expresión conjunta de los factores de transcripción ZFP238 y MYT1L, que mejoraron notablemente la producción de iN [125]. Por otro lado, la delección en células embrionarias de ratón de la proteína implicada en espermatogénesis MAGEB16, tiene como consecuencia la represión de *Tcfl5*, entre otros genes de espermatogénesis, tras sufrir una diferenciación al azar [227] [124]. Estos trabajos sugieren un posible papel de TCFL5 en la diferenciación hacia neuronas y/o línea germinal.

Las mESCs pueden dar lugar a los precursores de la línea germinal o precursores primordiales germinales (PGCs) bajo condiciones específicas de cultivo. Nuestros datos muestran que, tras la diferenciación a PGCs, la expresión de *Tcfl5* se vio inducida de igual manera que los marcadores de diferenciación. *Stella* se considera el marcador principal de PGCs [228] [229] y su sobreexpresión durante la diferenciación induce la expresión de marcadores tempranos de PGCs [230]. *Dazl* se considera también un marcador de PGC más tardío [230] y ratones *knockout* para este presentan una parada en la progresión de la línea germinal [231]. Durante la diferenciación a PGC, los valores de expresión de *Stella* y *Dazl* se vieron reducidos por la deficiencia de *Tcfl5*. Por otro lado, *Sohlh2* es un factor de transcripción bHLH altamente similar a *Tcfl5* en su secuencia. Su expresión es específica de testículo y oocito [232] y se ha descrito que ratones *knockout* para esta proteína son estériles por

problemas en la producción de gametos [233]. Aunque *Sohlh2* no se ha descrito en la diferenciación de mESCs a PGCs, sus similitudes con *Tcf5* nos llevaron a estudiar su expresión. *Sohlh2* se indujo durante la diferenciación al igual que *Stella*, *Dazl* y *Tcf5* sugiriendo un papel también en este proceso. La deficiencia en *Tcf5* también redujo los niveles de *Sohlh2*. Estos datos, aunque preliminares, sugieren un papel TCFL5 en la formación de PGCs donde su expresión es necesaria para su correcta formación y TCFL5 regularía la transcripción de *Dazl* y *Sohlh2*. Sin embargo, es necesario profundizar en el estudio de los marcadores y analizar el grado de diferenciación que se obtiene con la deficiencia de TCFL5.

Las PGCs son las únicas células que mantienen la expresión de los marcadores SOX2, NANOG y OCT4 tras el periodo de implantación y conservan, por tanto, propiedades de pluripotencia. SOX2 está expresado en PGCs y su falta de expresión produce una reducción dramática del número de precursores [234]. La deficiencia de OCT4 también produce una peor generación de PGCs [235]. La expresión de NANOG también se ha detectado en PGCs durante el desarrollo, pero no en células germinales de testículos y ovarios adultos [236]. Sin embargo, se han descrito funciones opuesta para SOX2 y NANOG. SOX2 reprime la formación *in vitro* de PGCs induciendo NANOG mientras que la expresión independiente de este por si solo induce la formación de PGCs [237]. Estos estudios demuestran la importancia de la expresión de estos factores de transcripción en la formación de PGCs y de la red transcripcional. Nuestros resultados muestran que hay una alteración de estos marcadores durante la diferenciación de mESCs a PGCs y que la deficiencia de *Tcf5* cambia el patrón de expresión sugiriendo un control de su expresión. No obstante, estudios en células pluripotentes derivadas de epiblasto (EpiSCs) mostraron una disminución significativa en los niveles de expresión de *Sox2*, *Nanog* y *Oct4* en células que darán lugar a PGCs [238]. Por tanto, el aumento de los marcadores de pluripotencia junto con una disminución de los marcadores de diferenciación a PGCs sugieren que la inducción de *Tcf5* es necesaria para la correcta diferenciación e indican que su deficiencia podría mantener las células en un estado más cercano a mESC que a PGC. En este sentido, es interesante mencionar que la expresión de TCFL5 aparece en las fases finales de diferenciación a línea germinal coincidente con una disminución de los genes de pluripotencia y que la deficiencia de TCFL5 hace que estos vuelvan a aumentar en esa fase final, confirmando un posible papel de TCFL5 en la regulación negativa de genes como *Nanog*, y *Sox2* observados en las células de cáncer de colon

Aunque no sabemos si estas alteraciones serán suficiente para impedir la futura formación de espermatozoides y es necesario un estudio más detallado [239], estos datos sugieren un papel importante de TCFL5 durante la diferenciación a línea germinal. Además, TCFL5 ya se había reportado como un gen implicado en espermatogénesis, aunque no en etapas tan tempranas [98]. TCFL5 controla transcripcionalmente la expresión de la chaperona *Clgn* [99]. Esta proteína funciona como chaperona de proteínas de superficie del espermatozoide implicadas en las interacciones de este con el oocito siendo los ratones deficientes para *Clgn* presentan problemas de fertilidad [240]. También se ha observado que TCFL5 interacciona con la proteína MEIG1 en espermatogénesis [100].

MEIG1 se ha descrito como esencial también en la fertilidad del ratón. Ratones deficientes en esta proteína fueron estériles como resultado de problemas en el estadio de elongación y condensación [241]. Por tanto, además de problemas de diferenciación en la línea germinal, la deficiencia de TCFL5 puede provocar una pérdida de expresión de CLGN e interacción con MEIG1 y por tanto producir problemas de fertilidad. Por otro lado, se ha visto una expresión consistente de *Tcfl5* en fibroblastos del estroma endometrial del útero tanto de humano como de ratón [242]. Estas células darán lugar a las células deciduales estromales necesarias para la implantación del embrión por lo que TCFL5 podría jugar un papel en este proceso de implantación.

La expresión de TCFL5 durante el proceso de desarrollo embrionario temprano junto con los datos mostrados en formación de gametos, hacen prever posibles problemas a la hora de obtener un ratón deficiente en la expresión de este gen. En los últimos años hemos generando una línea de ratones C57BL/6 condicionales para la deficiencia de la expresión de *Tcfl5* por delección de los exones e3 y e4 (*Tcfl5^{flox/flox}*). A día de hoy, no disponemos de ningún ratón homocigoto para la deficiencia de *Tcfl5* desde los primeros estadios del desarrollo sugiriendo problemas de fertilidad o desarrollo. Actualmente, nuestra estrategia se centra en la generación de ratones condicionales con la expresión de la recombinasa Cre bajo el control de tamoxifeno que nos permitirán estudiar *in vivo* estos procesos.

En resumen, hemos descrito que *TCFL5* es un gen con múltiples isoformas diferentes en ratón y humanos, lo que indica una evolución reciente del mismo. Así mismo, estas isoformas tienen papeles muy diferentes lo que complica mucho su análisis mediante experimentos genéticos. En humanos hemos visto que su deficiencia reduce drásticamente las propiedades tumorogénicas de las células tumorales de cáncer de colon. Reduce especialmente el crecimiento *in vivo* y la formación de MCTS. El efecto en el crecimiento *in vitro* no es tan importante, aunque afecta al ciclo celular. Esto sugiere la injerencia de TCFL5/CHA en otros procesos tumorales como interacción con el estroma, migración celular, EMT o CSCs. TCFL5/CHA reprimiría la expresión los marcadores de pluripotencia SOX2 y NANOG y también regularía a KLF4. Dada la importancia de estos 3 genes en los procesos de reprogramación celular implican a este factor en el control de la pluripotencia. En estos procesos TCFL5 y CHA parecen jugar papeles distintos siendo probablemente CHA la que tendría un papel más protumoral, activando las propiedades “*stem*” de la célula. La isoforma *TCFL5_E6* podría tener un papel adicional que complicaría más la interpretación de los datos mientras que la isoforma *TCFL5_E7* dada su escasa expresión podría no tener efecto en estos procesos. Así mismo, hemos visto que la expresión de TCFL5/CHA es fundamental para la inducción de la actividad NFkB2, regulando la expresión de genes anti-apoptóticos como *BCL2* siendo en este caso muy probablemente la isoforma TCFL5 la principal implicada en estos procesos mediante la unión a p100 NFkB (Figura 38). En ratón el efecto de TCFL5 es principalmente a nivel de diferenciación a línea germinal, probablemente reprimiendo la expresión de los genes de pluripotencia *Sox2* y *Nanog*, y controlando la expresión de genes importantes de este proceso como *Dazl* y *Sohlh2*.

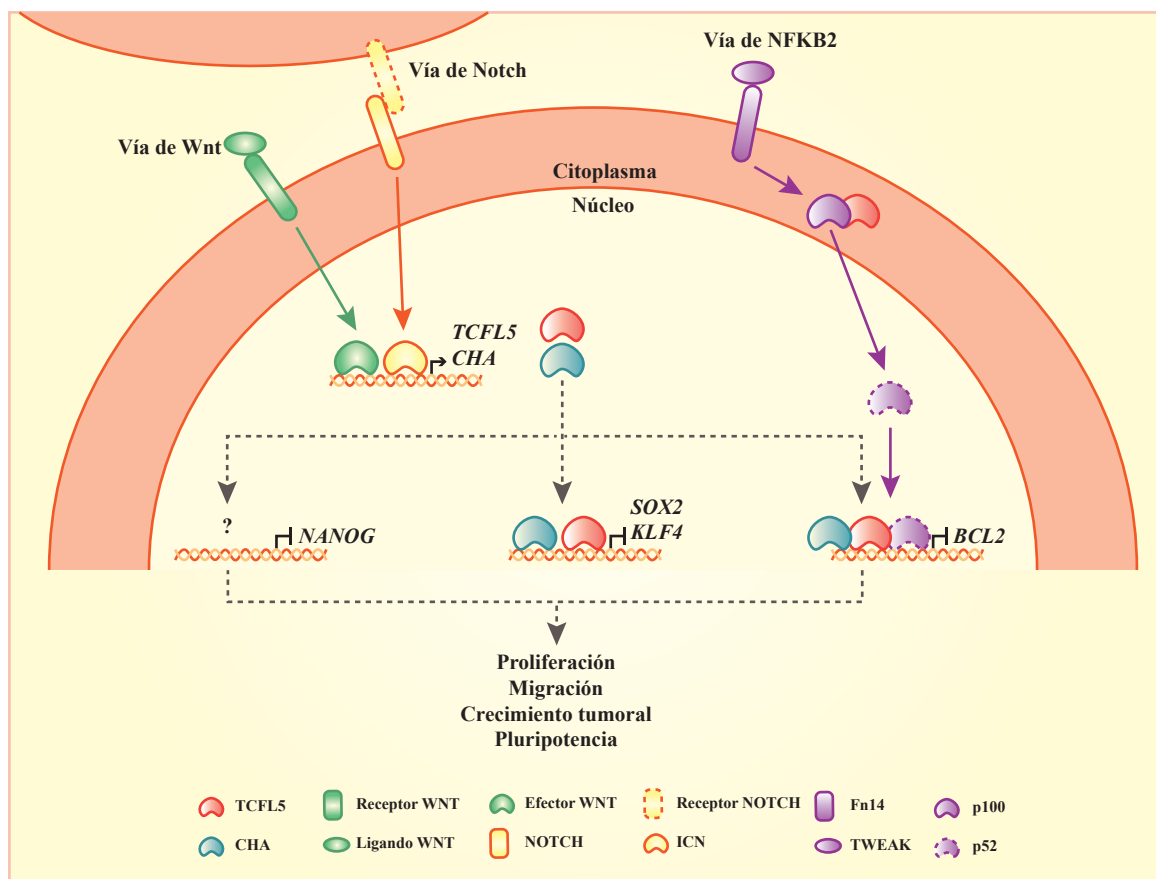


Figura 38. Esquema resumen de las vías de señalización implicadas en la expresión de *TCFL5/CHA* y de sus genes diana.

Conclusiones / Conclusions

Conclusiones

1. El locus de *TCFL5* es diferente en humanos y ratón. Su transcripción resulta en múltiples isoformas muy diferentes entre ambas especies. En humanos, la expresión es ubicua en todos los tejidos, siendo mayoritaria en testículo y cerebro, mientras que en ratón la expresión es específica de testículo, primeras etapas del desarrollo embrionario y células embrionarias.

2. La expresión de *TCFL5* es necesaria para el desarrollo del proceso tumoral de células de cáncer de colon. Su deficiencia inhibe la formación de MCTS, el crecimiento tumoral y afecta entre otros procesos a la proliferación. Esto, junto con su regulación por las vías de señalización de WNT y NOTCH, evidencian un papel fundamental en el desarrollo del cáncer colorrectal.

3. *TCFL5* y *CHA* controlan directamente la expresión de los marcadores de pluripotencia *SOX2*, *NANOG* y *KLF4* en células de cáncer de colon lo que les confiere un papel en el control de este proceso.

4. Las isoformas *TCFL5* y *CHA* controlan de manera opuesta la inducción de la actividad *NFKB2* por *TWEAK*. *TCFL5* y *CHA* controlan también la expresión del gen antiapoptótico *BCL2*.

5. Las isoformas de *TCFL5* se regulan entre ellas controlando directamente su transcripción. Esta regulación, junto con sus diferentes funciones a nivel celular y molecular, revelan un modelo altamente complejo en el que el ratio de expresión entre isoformas determina el fenotipo celular.

6. En ratón, *TCFL5* controla la diferenciación hacia precursores primordiales de la línea germinal reprimiendo los genes de pluripotencia *Sox2*, *Nanog* y *Oct4*, y controlando los genes de diferenciación *Dalz* y *Sohlh2*.

Conclusions

1. *TCFL5* loci differs in humans and mice. Its transcription results in multiple isoforms different between both species. In humans, its expression is ubiquitous in all tissues, principally testis and brain. However, in mice its expression is specific of testis, first stages of embryonic development and embryonic stem cells.

2. *TCFL5* expression is necessary for the tumoral development of colorectal cancer cells. Its complete deletion suppresses MCTS formation, tumor growth and other processes such as proliferation. This, along with its regulation for the WNT and NOTCH signaling pathways, reveals an essential role of *TCFL5* in colorectal cancer.

3. *TCFL5* and *CHA* directly control the expression of the pluripotency markers *SOX2*, *NANOG* and *KLF4* in colorectal cancer cells, which gives them an important role in this process.

4. *TCFL5* and *CHA* differentially regulate *NFKB2* activity induced by *TWEAK* and control the *BCL2* anti-apoptotic gene expression.

5. *TCFL5* isoforms directly regulate their own expression. This regulation, along with its different functions in the cells, reveals a highly complex model where the isoform expression ratio defines the cellular phenotype.

6. *TCFL5* controls primordial germinal cells differentiation in mice. In this process, *TCFL5* represses the pluripotency genes *Sox2*, *Nanog* and *Oct4* and it controls the differentiation genes *Dazl* and *Sohlh2*.

Bibliografia

Bibliografia

1. Cobb, M., *60 years ago, Francis Crick changed the logic of biology*. PLoS Biol, 2017. **15**(9): p. e2003243.
2. Levine, M. and R. Tjian, *Transcription regulation and animal diversity*. Nature, 2003. **424**(6945): p. 147-51.
3. Hughes, T.R., *Introduction to "a handbook of transcription factors"*. Subcell Biochem, 2011. **52**: p. 1-6.
4. Vaquerizas, J.M., et al., *A census of human transcription factors: function, expression and evolution*. Nat Rev Genet, 2009. **10**(4): p. 252-63.
5. Johnson, P.F. and S.L. McKnight, *Eukaryotic transcriptional regulatory proteins*. Annu Rev Biochem, 1989. **58**: p. 799-839.
6. Harrison, S.C., *A structural taxonomy of DNA-binding domains*. Nature, 1991. **353**(6346): p. 715-9.
7. Luscombe, N.M., et al., *An overview of the structures of protein-DNA complexes*. Genome Biol, 2000. **1**(1): p. REVIEWS001.
8. Brennan, R.G. and B.W. Matthews, *The helix-turn-helix DNA binding motif*. J Biol Chem, 1989. **264**(4): p. 1903-6.
9. Murre, C., et al., *Structure and function of helix-loop-helix proteins*. Biochim Biophys Acta, 1994. **1218**(2): p. 129-35.
10. Atchley, W.R. and W.M. Fitch, *A natural classification of the basic helix-loop-helix class of transcription factors*. Proc Natl Acad Sci U S A, 1997. **94**(10): p. 5172-6.
11. Ledent, V., O. Paquet, and M. Vervoort, *Phylogenetic analysis of the human basic helix-loop-helix proteins*. Genome Biol, 2002. **3**(6): p. RESEARCH0030.
12. Jones, S., *An overview of the basic helix-loop-helix proteins*. Genome Biol, 2004. **5**(6): p. 226.
13. Massari, M.E. and C. Murre, *Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms*. Mol Cell Biol, 2000. **20**(2): p. 429-40.
14. Krivega, I. and A. Dean, *Enhancer and promoter interactions-long distance calls*. Curr Opin Genet Dev, 2012. **22**(2): p. 79-85.
15. Dikstein, R., *The unexpected traits associated with core promoter elements*. Transcription, 2011. **2**(5): p. 201-6.
16. Latchman, D.S., *Transcription factors: an overview*. Int J Exp Pathol, 1993. **74**(5): p. 417-22.
17. Calkhoven, C.F. and G. Ab, *Multiple steps in the regulation of transcription-factor level and activity*. Biochem J, 1996. **317** (Pt 2): p. 329-42.
18. Filtz, T.M., W.K. Vogel, and M. Leid, *Regulation of transcription factor activity by interconnected post-translational modifications*. Trends Pharmacol Sci, 2014. **35**(2): p. 76-85.
19. Reiter, F., S. Wienerroither, and A. Stark, *Combinatorial function of transcription factors and cofactors*. Curr Opin Genet Dev, 2017. **43**: p. 73-81.
20. Amoutzias, G.D., et al., *Choose your partners: dimerization in eukaryotic transcription factors*. Trends Biochem Sci, 2008. **33**(5): p. 220-9.
21. Grandori, C., et al., *The Myc/Max/Mad network and the transcriptional control of cell behavior*. Annu Rev Cell Dev Biol, 2000. **16**: p. 653-99.

22. Luscher, B., *Function and regulation of the transcription factors of the Myc/Max/Mad network*. Gene, 2001. **277**(1-2): p. 1-14.
23. Lluís, F., et al., *E47 phosphorylation by p38 MAPK promotes MyoD/E47 association and muscle-specific gene transcription*. EMBO J, 2005. **24**(5): p. 974-84.
24. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries*. CA Cancer J Clin, 2018. **68**(6): p. 394-424.
25. Asociación de Oncología Médica, *Las cifras del cáncer de España 2018*. 2018.
26. National Center Institute, *What is Cancer?* 2015.
27. Stratton, M.R., P.J. Campbell, and P.A. Futreal, *The cancer genome*. Nature, 2009. **458**(7239): p. 719-24.
28. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation*. Cell, 2011. **144**(5): p. 646-74.
29. Bradner, J.E., D. Hnisz, and R.A. Young, *Transcriptional Addiction in Cancer*. Cell, 2017. **168**(4): p. 629-643.
30. Morris, S.A. and G.Q. Daley, *A blueprint for engineering cell fate: current technologies to reprogram cell identity*. Cell Res, 2013. **23**(1): p. 33-48.
31. Sanda, T., et al., *Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia*. Cancer Cell, 2012. **22**(2): p. 209-21.
32. Lin, C.Y., et al., *Transcriptional amplification in tumor cells with elevated c-Myc*. Cell, 2012. **151**(1): p. 56-67.
33. Nie, Z., et al., *c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells*. Cell, 2012. **151**(1): p. 68-79.
34. Gabay, M., Y. Li, and D.W. Felsher, *MYC activation is a hallmark of cancer initiation and maintenance*. Cold Spring Harb Perspect Med, 2014. **4**(6).
35. Darnell, J.E., Jr., *Transcription factors as targets for cancer therapy*. Nat Rev Cancer, 2002. **2**(10): p. 740-9.
36. Dalla-Favera, R., et al., *Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells*. Proc Natl Acad Sci U S A, 1982. **79**(24): p. 7824-7.
37. Taub, R., et al., *Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells*. Proc Natl Acad Sci U S A, 1982. **79**(24): p. 7837-41.
38. Erikson, J., et al., *Transcriptional activation of the translocated c-myc oncogene in burkitt lymphoma*. Proc Natl Acad Sci U S A, 1983. **80**(3): p. 820-4.
39. Petitjean, A., et al., *Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database*. Hum Mutat, 2007. **28**(6): p. 622-9.
40. Goh, A.M., C.R. Coffill, and D.P. Lane, *The role of mutant p53 in human cancer*. J Pathol, 2011. **223**(2): p. 116-26.
41. Aurelio, O.N., et al., *p53 mutants have selective dominant-negative effects on apoptosis but not*

- growth arrest in human cancer cell lines*. Mol Cell Biol, 2000. **20**(3): p. 770-8.
42. Yan, C. and P.J. Higgins, *Drugging the undruggable: transcription therapy for cancer*. Biochim Biophys Acta, 2013. **1835**(1): p. 76-85.
 43. Lambert, M., et al., *Targeting Transcription Factors for Cancer Treatment*. Molecules, 2018. **23**(6).
 44. Siegel, R.L., et al., *Colorectal cancer statistics, 2017*. CA Cancer J Clin, 2017. **67**(3): p. 177-193.
 45. Asociación Española Contra el Cáncer, *Incidencia y mortalidad de cáncer colorrectal en España en la población entre 50 y 69 años*. 2019.
 46. Edwards, B.K., et al., *Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates*. Cancer, 2010. **116**(3): p. 544-73.
 47. Geng, F., et al., *Molecular Targeted Drugs and Treatment of Colorectal Cancer: Recent Progress and Future Perspectives*. Cancer Biother Radiopharm, 2017. **32**(5): p. 149-160.
 48. Kikuchi, Y., W.N. Dinjens, and F.T. Bosman, *Proliferation and apoptosis in proliferative lesions of the colon and rectum*. Virchows Arch, 1997. **431**(2): p. 111-7.
 49. Grady, W.M., *Genetic testing for high-risk colon cancer patients*. Gastroenterology, 2003. **124**(6): p. 1574-94.
 50. Sancho, E., E. Batlle, and H. Clevers, *Signaling pathways in intestinal development and cancer*. Annu Rev Cell Dev Biol, 2004. **20**: p. 695-723.
 51. Moran, A., et al., *Differential colorectal carcinogenesis: Molecular basis and clinical relevance*. World J Gastrointest Oncol, 2010. **2**(3): p. 151-8.
 52. Fearon, E.R. and B. Vogelstein, *A genetic model for colorectal tumorigenesis*. Cell, 1990. **61**(5): p. 759-67.
 53. Arends, J.W., *Molecular interactions in the Vogelstein model of colorectal carcinoma*. J Pathol, 2000. **190**(4): p. 412-6.
 54. Cattaneo, E., et al., *Preinvasive colorectal lesion transcriptomes correlate with endoscopic morphology (polypoid vs. nonpolypoid)*. EMBO Mol Med, 2011. **3**(6): p. 334-47.
 55. Maglietta, R., et al., *Molecular pathways undergoing dramatic transcriptomic changes during tumor development in the human colon*. BMC Cancer, 2012. **12**: p. 608.
 56. Vonlanthen, J., et al., *A comprehensive look at transcription factor gene expression changes in colorectal adenomas*. BMC Cancer, 2014. **14**: p. 46.
 57. Zhou, B. and R. Guo, *Genomic and regulatory characteristics of significant transcription factors in colorectal cancer metastasis*. Sci Rep, 2018. **8**(1): p. 17836.
 58. Aiken, C.E., et al., *The direct measurement of embryogenic volume and nucleo-cytoplasmic ratio during mouse pre-implantation development*. Reproduction, 2004. **128**(5): p. 527-35.
 59. Kelly, S.J., *Studies of the developmental potential of 4- and 8-cell stage mouse blastomeres*. J Exp Zool, 1977. **200**(3): p. 365-76.
 60. Mihajlovic, A.I. and A.W. Bruce, *The first cell-fate decision of mouse preimplantation embryo development: integrating cell position and polarity*. Open Biol, 2017. **7**(11).
 61. Svoboda, P., *Mammalian zygotic genome activation*. Semin Cell Dev Biol, 2018. **84**: p. 118-126.

62. Godini, R. and H. Fallahi, *Dynamics changes in the transcription factors during early human embryonic development*. J Cell Physiol, 2019. **234**(5): p. 6489-6502.
63. Niakan, K.K., et al., *Human pre-implantation embryo development*. Development, 2012. **139**(5): p. 829-41.
64. Xie, D., et al., *Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species*. Genome Res, 2010. **20**(6): p. 804-15.
65. Gabdoulline, R., et al., *Differences in the Early Development of Human and Mouse Embryonic Stem Cells*. PLoS One, 2015. **10**(10): p. e0140803.
66. Plachta, N., et al., *Oct4 kinetics predict cell lineage patterning in the early mammalian embryo*. Nat Cell Biol, 2011. **13**(2): p. 117-23.
67. Kaur, G., et al., *Probing transcription factor diffusion dynamics in the living mammalian embryo with photoactivatable fluorescence correlation spectroscopy*. Nat Commun, 2013. **4**: p. 1637.
68. Niwa, H., et al., *Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation*. Cell, 2005. **123**(5): p. 917-29.
69. Strumpf, D., et al., *Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst*. Development, 2005. **132**(9): p. 2093-102.
70. Chazaud, C., et al., *Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway*. Dev Cell, 2006. **10**(5): p. 615-24.
71. Xenopoulos, P., et al., *Heterogeneities in Nanog Expression Drive Stable Commitment to Pluripotency in the Mouse Blastocyst*. Cell Rep, 2015. **10**(9): p. 1508-1520.
72. Singh, A.M., et al., *A heterogeneous expression pattern for Nanog in embryonic stem cells*. Stem Cells, 2007. **25**(10): p. 2534-42.
73. Avilion, A.A., et al., *Multipotent cell lineages in early mouse development depend on SOX2 function*. Genes Dev, 2003. **17**(1): p. 126-40.
74. De Los Angeles, A., et al., *Hallmarks of pluripotency*. Nature, 2015. **525**(7570): p. 469-78.
75. Smith, A.G., *Embryo-derived stem cells: of mice and men*. Annu Rev Cell Dev Biol, 2001. **17**: p. 435-62.
76. Boyer, L.A., et al., *Core transcriptional regulatory circuitry in human embryonic stem cells*. Cell, 2005. **122**(6): p. 947-56.
77. Rizzino, A., *Concise review: The Sox2-Oct4 connection: critical players in a much larger interdependent network integrated at multiple levels*. Stem Cells, 2013. **31**(6): p. 1033-9.
78. Kuroda, T., et al., *Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression*. Mol Cell Biol, 2005. **25**(6): p. 2475-85.
79. Rodda, D.J., et al., *Transcriptional regulation of nanog by OCT4 and SOX2*. J Biol Chem, 2005. **280**(26): p. 24731-7.
80. Okumura-Nakanishi, S., et al., *Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells*. J Biol Chem, 2005. **280**(7): p. 5307-17.
81. Masui, S., et al., *Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells*. Nat Cell Biol, 2007. **9**(6): p. 625-35.
82. Niwa, H., J. Miyazaki, and A.G. Smith, *Quantitative expression of Oct-3/4 defines differentiation,*

- dedifferentiation or self-renewal of ES cells*. Nat Genet, 2000. **24**(4): p. 372-6.
83. Kopp, J.L., et al., *Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells*. Stem Cells, 2008. **26**(4): p. 903-11.
 84. Niwa, H., et al., *A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells*. Nature, 2009. **460**(7251): p. 118-22.
 85. Ying, Q.L., et al., *BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3*. Cell, 2003. **115**(3): p. 281-92.
 86. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors*. Cell, 2006. **126**(4): p. 663-76.
 87. Liu, X., et al., *Yamanaka factors critically regulate the developmental signaling network in mouse embryonic stem cells*. Cell Res, 2008. **18**(12): p. 1177-89.
 88. Papapetrou, E.P., *Induced pluripotent stem cells, past and future*. Science, 2016. **353**(6303): p. 991-992.
 89. Stastna, M. and J.E. Van Eyk, *Analysis of protein isoforms: can we do it better?* Proteomics, 2012. **12**(19-20): p. 2937-48.
 90. Wang, E.T., et al., *Alternative isoform regulation in human tissue transcriptomes*. Nature, 2008. **456**(7221): p. 470-6.
 91. Blencowe, B.J., *Alternative splicing: new insights from global analyses*. Cell, 2006. **126**(1): p. 37-47.
 92. Gonzalez-Porta, M., et al., *Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene*. Genome Biol, 2013. **14**(7): p. R70.
 93. Uhlen, M., et al., *Proteomics. Tissue-based map of the human proteome*. Science, 2015. **347**(6220): p. 1260419.
 94. Scotti, M.M. and M.S. Swanson, *RNA mis-splicing in disease*. Nat Rev Genet, 2016. **17**(1): p. 19-32.
 95. Baralle, F.E. and J. Giudice, *Alternative splicing as a regulator of development and tissue identity*. Nat Rev Mol Cell Biol, 2017. **18**(7): p. 437-451.
 96. Zhao, Q., et al., *Tumor-specific isoform switch of the fibroblast growth factor receptor 2 underlies the mesenchymal and malignant phenotypes of clear cell renal cell carcinomas*. Clin Cancer Res, 2013. **19**(9): p. 2460-72.
 97. El Marabti, E. and I. Younis, *The Cancer Spliceome: Reprogramming of Alternative Splicing in Cancer*. Front Mol Biosci, 2018. **5**: p. 80.
 98. Maruyama, O., et al., *Cloning of TCFL5 encoding a novel human basic helix-loop-helix motif protein that is specifically expressed in primary spermatocytes at the pachytene stage*. Cytogenet Cell Genet, 1998. **82**(1-2): p. 41-5.
 99. Siep, M., et al., *Basic helix-loop-helix transcription factor Tcf15 interacts with the Calmegin gene promoter in mouse spermatogenesis*. Nucleic Acids Res, 2004. **32**(21): p. 6425-36.
 100. Shi, Y., et al., *The mouse transcription factor-like 5 gene encodes a protein localized in the manchette and centriole of the elongating spermatid*. Andrology, 2013. **1**(3): p. 431-9.
 101. Girones, N., et al., *Dominant T- and B-cell epitopes in an autoantigen linked to Chagas' disease*. J Clin Invest, 2001. **107**(8): p. 985-93.

102. Girones, N., et al., *Antibodies to an epitope from the Cha human autoantigen are markers of Chagas' disease*. Clin Diagn Lab Immunol, 2001. **8**(6): p. 1039-43.
103. Rodriguez, C.I., N. Girones, and M. Fresno, *Cha, a basic helix-loop-helix transcription factor involved in the regulation of upstream stimulatory factor activity*. J Biol Chem, 2003. **278**(44): p. 43135-45.
104. Weerkamp, F., et al., *Identification of Notch target genes in uncommitted T-cell progenitors: No direct induction of a T-cell specific gene program*. Leukemia, 2006. **20**(11): p. 1967-77.
105. Kopan, R., *Notch signaling*. Cold Spring Harb Perspect Biol, 2012. **4**(10).
106. Hanten, J.A., et al., *Comparison of human B cell activation by TLR7 and TLR9 agonists*. BMC Immunol, 2008. **9**: p. 39.
107. Tomaru, Y., et al., *Regulatory interdependence of myeloid transcription factors revealed by Matrix RNAi analysis*. Genome Biol, 2009. **10**(11): p. R121.
108. Smith, A.J., et al., *Host genes associated with HIV-1 replication in lymphatic tissue*. J Immunol, 2010. **185**(9): p. 5417-24.
109. Sikdar, S. and S. Datta, *A novel statistical approach for identification of the master regulator transcription factor*. BMC Bioinformatics, 2017. **18**(1): p. 79.
110. Bayles, R.G., et al., *Transcriptomic and neurochemical analysis of the stellate ganglia in mice highlights sex differences*. Sci Rep, 2018. **8**(1): p. 8963.
111. Dardousis, K., et al., *Identification of differentially expressed genes involved in the formation of multicellular tumor spheroids by HT-29 colon carcinoma cells*. Mol Ther, 2007. **15**(1): p. 94-102.
112. Gottfried, E., et al., *Brave little world: spheroids as an in vitro model to study tumor-immune-cell interactions*. Cell Cycle, 2006. **5**(7): p. 691-5.
113. LaBarbera, D.V., B.G. Reid, and B.H. Yoo, *The multicellular tumor spheroid model for high-throughput cancer drug discovery*. Expert Opin Drug Discov, 2012. **7**(9): p. 819-30.
114. Huang, B.W. and J.Q. Gao, *Application of 3D cultured multicellular spheroid tumor models in tumor-targeted drug delivery system research*. J Control Release, 2018. **270**: p. 246-259.
115. Sillars-Hardebol, A.H., et al., *TPX2 and AURKA promote 20q amplicon-driven colorectal adenoma to carcinoma progression*. Gut, 2012. **61**(11): p. 1568-75.
116. Carvalho, B., et al., *Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression*. Gut, 2009. **58**(1): p. 79-89.
117. Holleman, A., et al., *Gene-expression patterns in drug-resistant acute lymphoblastic leukemia cells and response to treatment*. N Engl J Med, 2004. **351**(6): p. 533-42.
118. Silveira, V.S., et al., *Gene expression pattern contributing to prognostic factors in childhood acute lymphoblastic leukemia*. Leuk Lymphoma, 2013. **54**(2): p. 310-4.
119. Gandemer, V., et al., *Five distinct biological processes and 14 differentially expressed genes characterize TEL/AML1-positive leukemia*. BMC Genomics, 2007. **8**: p. 385.
120. Busche, S., et al., *Integration of high-resolution methylome and transcriptome analyses to dissect epigenomic changes in childhood acute lymphoblastic leukemia*. Cancer Res, 2013. **73**(14): p. 4323-36.
121. Looijenga, L.H., et al., *Genomic and expression profiling of human spermatocytic seminomas: primary*

- spermatocyte as tumorigenic precursor and DMRT1 as candidate chromosome 9 gene*. *Cancer Res*, 2006. **66**(1): p. 290-302.
122. Chen, G., et al., *Single-cell analyses of X Chromosome inactivation dynamics and pluripotency during differentiation*. *Genome Res*, 2016. **26**(10): p. 1342-1354.
 123. Foygel, K., et al., *A novel and critical role for Oct4 as a regulator of the maternal-embryonic transition*. *PLoS One*, 2008. **3**(12): p. e4109.
 124. Gaspar, J.A., et al., *Depletion of Mageb16 induces differentiation of pluripotent stem cells predominantly into mesodermal derivatives*. *Sci Rep*, 2017. **7**(1): p. 14285.
 125. Wapinski, O.L., et al., *Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons*. *Cell*, 2013. **155**(3): p. 621-35.
 126. Arbab, M. and R.I. Sherwood, *Self-Cloning CRISPR*. *Curr Protoc Stem Cell Biol*, 2016. **38**: p. 5B 5 1-5B 5 16.
 127. Consortium, G.T., *The Genotype-Tissue Expression (GTEx) project*. *Nat Genet*, 2013. **45**(6): p. 580-5.
 128. Ernst, J. and M. Kellis, *Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues*. *Nat Biotechnol*, 2015. **33**(4): p. 364-76.
 129. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome*. *Nature*, 2012. **489**(7414): p. 57-74.
 130. Lizio, M., et al., *Gateways to the FANTOM5 promoter level mammalian expression atlas*. *Genome Biol*, 2015. **16**: p. 22.
 131. Rhodes, D.R., et al., *ONCOMINE: a cancer microarray database and integrated data-mining platform*. *Neoplasia*, 2004. **6**(1): p. 1-6.
 132. Skrzypczak, M., et al., *Modeling oncogenic signaling in colon tumors by multidirectional analyses of microarray data directed for maximization of analytical reliability*. *PLoS One*, 2010. **5**(10).
 133. Hong, Y., et al., *A 'metastasis-prone' signature for early-stage mismatch-repair proficient sporadic colorectal cancer patients and its implications for possible therapeutics*. *Clin Exp Metastasis*, 2010. **27**(2): p. 83-90.
 134. Cancer Genome Atlas, N., *Comprehensive molecular characterization of human colon and rectal cancer*. *Nature*, 2012. **487**(7407): p. 330-7.
 135. Gao, J., et al., *Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal*. *Sci Signal*, 2013. **6**(269): p. p11.
 136. Abukhdeir, A.M. and B.H. Park, *P21 and p27: roles in carcinogenesis and drug resistance*. *Expert Rev Mol Med*, 2008. **10**: p. e19.
 137. Yan, J., et al., *Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites*. *Cell*, 2013. **154**(4): p. 801-13.
 138. Frenzel, A., et al., *Bcl2 family proteins in carcinogenesis and the treatment of cancer*. *Apoptosis*, 2009. **14**(4): p. 584-96.
 139. Viatour, P., et al., *NF- kappa B2/p100 induces Bcl-2 expression*. *Leukemia*, 2003. **17**(7): p. 1349-56.
 140. Saitoh, T., et al., *TWEAK induces NF-kappaB2 p100 processing and long lasting NF-kappaB activation*. *J Biol Chem*, 2003. **278**(38): p. 36005-12.

141. Nakayama, M., et al., *Multiple pathways of TWEAK-induced cell death*. J Immunol, 2002. **168**(2): p. 734-43.
142. Nakayama, M., et al., *Fibroblast growth factor-inducible 14 mediates multiple pathways of TWEAK-induced cell death*. J Immunol, 2003. **170**(1): p. 341-8.
143. Ishiguro, T., et al., *Tumor-derived spheroids: Relevance to cancer stem cells and clinical applications*. Cancer Sci, 2017. **108**(3): p. 283-289.
144. Dalerba, P., et al., *Phenotypic characterization of human colorectal cancer stem cells*. Proc Natl Acad Sci U S A, 2007. **104**(24): p. 10158-63.
145. Uchida, H., et al., *Overexpression of leucine-rich repeat-containing G protein-coupled receptor 5 in colorectal cancer*. Cancer Sci, 2010. **101**(7): p. 1731-7.
146. Merlos-Suarez, A., et al., *The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse*. Cell Stem Cell, 2011. **8**(5): p. 511-24.
147. Ricci-Vitiani, L., et al., *Identification and expansion of human colon-cancer-initiating cells*. Nature, 2007. **445**(7123): p. 111-5.
148. Yu, J., et al., *Human induced pluripotent stem cells free of vector and transgene sequences*. Science, 2009. **324**(5928): p. 797-801.
149. Leis, O., et al., *Sox2 expression in breast tumours and activation in breast cancer stem cells*. Oncogene, 2012. **31**(11): p. 1354-65.
150. Kelemen, O., et al., *Function of alternative splicing*. Gene, 2013. **514**(1): p. 1-30.
151. Azuara, V., et al., *Chromatin signatures of pluripotent cell lines*. Nat Cell Biol, 2006. **8**(5): p. 532-8.
152. Bernstein, B.E., et al., *A bivalent chromatin structure marks key developmental genes in embryonic stem cells*. Cell, 2006. **125**(2): p. 315-26.
153. Mochizuki, K., et al., *Implication of DNA demethylation and bivalent histone modification for selective gene regulation in mouse primordial germ cells*. PLoS One, 2012. **7**(9): p. e46036.
154. Erkek, S., et al., *Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa*. Nat Struct Mol Biol, 2013. **20**(7): p. 868-75.
155. Pei, B., et al., *The GENCODE pseudogene resource*. Genome Biol, 2012. **13**(9): p. R51.
156. Zarzour, P., et al., *Single nucleotide polymorphism array profiling identifies distinct chromosomal aberration patterns across colorectal adenomas and carcinomas*. Genes Chromosomes Cancer, 2015. **54**(5): p. 303-14.
157. Ahmed, D., et al., *Epigenetic and genetic features of 24 colon cancer cell lines*. Oncogenesis, 2013. **2**: p. e71.
158. Strzalka, W. and A. Ziemienowicz, *Proliferating cell nuclear antigen (PCNA): a key factor in DNA replication and cell cycle regulation*. Ann Bot, 2011. **107**(7): p. 1127-40.
159. Waga, S., et al., *The p21 inhibitor of cyclin-dependent kinases controls DNA replication by interaction with PCNA*. Nature, 1994. **369**(6481): p. 574-8.
160. Rousseau, D., et al., *Growth inhibition by CDK-cyclin and PCNA binding domains of p21 occurs by distinct mechanisms and is regulated by ubiquitin-proteasome pathway*. Oncogene, 1999. **18**(21): p. 3290-302.

161. Chu, G. and W. Yang, *Here comes the sun: recognition of UV-damaged DNA*. Cell, 2008. **135**(7): p. 1172-4.
162. Vermeulen, L., et al., *Single-cell cloning of colon cancer stem cells reveals a multi-lineage differentiation capacity*. Proc Natl Acad Sci U S A, 2008. **105**(36): p. 13427-32.
163. Kabra, N., et al., *SirT1 is an inhibitor of proliferation and tumor formation in colon cancer*. J Biol Chem, 2009. **284**(27): p. 18210-7.
164. Yu, J., et al., *CDX2 inhibits the proliferation and tumor formation of colon cancer cells by suppressing Wnt/beta-catenin signaling via transactivation of GSK-3beta and Axin2 expression*. Cell Death Dis, 2019. **10**(1): p. 26.
165. Scheel, C. and R.A. Weinberg, *Cancer stem cells and epithelial-mesenchymal transition: concepts and molecular links*. Semin Cancer Biol, 2012. **22**(5-6): p. 396-403.
166. Saitoh, M., *Involvement of partial EMT in cancer progression*. J Biochem, 2018. **164**(4): p. 257-264.
167. Okegawa, T., et al., *The role of cell adhesion molecule in cancer progression and its application in cancer therapy*. Acta Biochim Pol, 2004. **51**(2): p. 445-57.
168. Chung Moh, M., L. Hoon Lee, and S. Shen, *Cloning and characterization of hepaCAM, a novel Ig-like cell adhesion molecule suppressed in human hepatocellular carcinoma*. J Hepatol, 2005. **42**(6): p. 833-41.
169. Liu, W., et al., *CEACAM1 impedes thyroid cancer growth but promotes invasiveness: a putative mechanism for early metastases*. Oncogene, 2007. **26**(19): p. 2747-58.
170. Pinson, J., et al., *Positive autoregulation of the transcription factor Pax6 in response to increased levels of either of its major isoforms, Pax6 or Pax6(5a), in cultured cells*. BMC Dev Biol, 2006. **6**: p. 25.
171. Luedde, T., et al., *C/EBP beta isoforms LIP and LAP modulate progression of the cell cycle in the regenerating mouse liver*. Hepatology, 2004. **40**(2): p. 356-65.
172. Yang, X., et al., *Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing*. Cell, 2016. **164**(4): p. 805-17.
173. Tegowski, M. and A. Baldwin, *Noncanonical NF-kappaB in Cancer*. Biomedicines, 2018. **6**(2).
174. Ge, Q.L., et al., *RelB/NF-kappaB links cell cycle transition and apoptosis to endometrioid adenocarcinoma tumorigenesis*. Cell Death Dis, 2016. **7**(10): p. e2402.
175. Wang, Y., et al., *Tumor-suppressor NFkappaB2 p100 interacts with ERK2 and stabilizes PTEN mRNA via inhibition of miR-494*. Oncogene, 2016. **35**(31): p. 4080-90.
176. Jacque, E., et al., *RelB inhibits cell proliferation and tumor growth through p53 transcriptional activation*. Oncogene, 2013. **32**(21): p. 2661-9.
177. Winkles, J.A., *The TWEAK-Fn14 cytokine-receptor axis: discovery, biology and therapeutic targeting*. Nat Rev Drug Discov, 2008. **7**(5): p. 411-25.
178. Kawakita, T., et al., *Functional expression of TWEAK in human colonic adenocarcinoma cells*. Int J Oncol, 2005. **26**(1): p. 87-93.
179. Chicheportiche, Y., et al., *TWEAK, a new secreted ligand in the tumor necrosis factor family that weakly induces apoptosis*. J Biol Chem, 1997. **272**(51): p. 32401-10.
180. Lin, B.R., et al., *Prognostic significance of TWEAK expression in colorectal cancer and effect of its*

- inhibition on invasion*. Ann Surg Oncol, 2012. **19 Suppl 3**: p. S385-94.
181. Culp, P.A., et al., *Antibodies to TWEAK receptor inhibit human tumor growth through dual mechanisms*. Clin Cancer Res, 2010. **16**(2): p. 497-508.
 182. Armstrong, C.L., et al., *TWEAK activation of the non-canonical NF-kappaB signaling pathway differentially regulates melanoma and prostate cancer cell invasion*. Oncotarget, 2016. **7**(49): p. 81474-81492.
 183. Ok, C.Y., et al., *Evaluation of NF-kappaB subunit expression and signaling pathway activation demonstrates that p52 expression confers better outcome in germinal center B-cell-like diffuse large B-cell lymphoma in association with CD30 and BCL2 functions*. Mod Pathol, 2015. **28**(9): p. 1202-13.
 184. Ilyas, M., et al., *Loss of Bcl-2 expression correlates with tumour recurrence in colorectal cancer*. Gut, 1998. **43**(3): p. 383-7.
 185. Letai, A., et al., *Antiapoptotic BCL-2 is required for maintenance of a model leukemia*. Cancer Cell, 2004. **6**(3): p. 241-9.
 186. Korinek, V., et al., *Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma*. Science, 1997. **275**(5307): p. 1784-7.
 187. Matano, M., et al., *Modeling colorectal cancer using CRISPR-Cas9-mediated engineering of human intestinal organoids*. Nat Med, 2015. **21**(3): p. 256-62.
 188. Xie, J., et al., *Inhibition of Tcf-4 induces apoptosis and enhances chemosensitivity of colon cancer cells*. PLoS One, 2012. **7**(9): p. e45617.
 189. Shin, S.H., et al., *A Small Molecule Inhibitor of the beta-Catenin-TCF4 Interaction Suppresses Colorectal Cancer Growth In Vitro and In Vivo*. EBioMedicine, 2017. **25**: p. 22-31.
 190. Hrckulak, D., et al., *Wnt Effector TCF4 Is Dispensable for Wnt Signaling in Human Cancer Cells*. Genes (Basel), 2018. **9**(9).
 191. Chu, D., et al., *Notch1 and Notch2 have opposite prognostic effects on patients with colorectal cancer*. Ann Oncol, 2011. **22**(11): p. 2440-7.
 192. Zhang, Y., et al., *Notch1 regulates the growth of human colon cancers*. Cancer, 2010. **116**(22): p. 5207-18.
 193. Sikandar, S.S., et al., *NOTCH signaling is required for formation and self-renewal of tumor-initiating cells and for repression of secretory cell differentiation in colon cancer*. Cancer Res, 2010. **70**(4): p. 1469-78.
 194. Vishnubalaji, R., et al., *Molecular profiling of ALDH1(+) colorectal cancer stem cells reveals preferential activation of MAPK, FAK, and oxidative stress pro-survival signalling pathways*. Oncotarget, 2018. **9**(17): p. 13551-13564.
 195. Martinez-Cruzado, L., et al., *Aldh1 Expression and Activity Increase During Tumor Evolution in Sarcoma Cancer Stem Cell Populations*. Sci Rep, 2016. **6**: p. 27878.
 196. Hirsch, D., et al., *LGR5 positivity defines stem-like cells in colorectal cancer*. Carcinogenesis, 2014. **35**(4): p. 849-58.
 197. Liu, W., et al., *LGR5 promotes epithelial ovarian cancer proliferation, metastasis, and epithelial-mesenchymal transition through the Notch1 signaling pathway*. Cancer Med, 2018.

198. O'Brien, C.A., et al., *A human colon cancer cell capable of initiating tumour growth in immunodeficient mice*. *Nature*, 2007. **445**(7123): p. 106-10.
199. Batlle, E., et al., *Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB*. *Cell*, 2002. **111**(2): p. 251-63.
200. Batlle, E., et al., *EphB receptor activity suppresses colorectal cancer progression*. *Nature*, 2005. **435**(7045): p. 1126-30.
201. Munro, M.J., et al., *Cancer stem cells in colorectal cancer: a review*. *J Clin Pathol*, 2018. **71**(2): p. 110-116.
202. Maurizi, G., et al., *Sox2 is required for tumor development and cancer cell proliferation in osteosarcoma*. *Oncogene*, 2018. **37**(33): p. 4626-4632.
203. Lundberg, I.V., et al., *SOX2 expression is associated with a cancer stem cell state and down-regulation of CDX2 in colorectal cancer*. *BMC Cancer*, 2016. **16**: p. 471.
204. Takeda, K., et al., *Sox2 is associated with cancer stem-like properties in colorectal cancer*. *Sci Rep*, 2018. **8**(1): p. 17639.
205. Lundberg, I.V., et al., *SOX2 expression is regulated by BRAF and contributes to poor patient prognosis in colorectal cancer*. *PLoS One*, 2014. **9**(7): p. e101957.
206. Zheng, J., et al., *Sox2 modulates motility and enhances progression of colorectal cancer via the Rho-ROCK signaling pathway*. *Oncotarget*, 2017. **8**(58): p. 98635-98645.
207. Saigusa, S., et al., *Correlation of CD133, OCT4, and SOX2 in rectal cancer and their association with distant recurrence after chemoradiotherapy*. *Ann Surg Oncol*, 2009. **16**(12): p. 3488-98.
208. Neumann, J., et al., *SOX2 expression correlates with lymph-node metastases and distant spread in right-sided colon cancer*. *BMC Cancer*, 2011. **11**: p. 518.
209. Amini, S., et al., *The expressions of stem cell markers: Oct4, Nanog, Sox2, nucleostemin, Bmi, Zfx, Tcl1, Tbx3, Dppa4, and Esrrb in bladder, colon, and prostate cancer, and certain cancer cell lines*. *Anat Cell Biol*, 2014. **47**(1): p. 1-11.
210. Zbinden, M., et al., *NANOG regulates glioma stem cells and is essential in vivo acting in a cross-functional network with GLI1 and p53*. *EMBO J*, 2010. **29**(15): p. 2659-74.
211. Jeter, C.R., et al., *NANOG promotes cancer stem cell characteristics and prostate cancer resistance to androgen deprivation*. *Oncogene*, 2011. **30**(36): p. 3833-45.
212. Ibrahim, E.E., et al., *Embryonic NANOG activity defines colorectal cancer stem cells and modulates through AP1- and TCF-dependent mechanisms*. *Stem Cells*, 2012. **30**(10): p. 2076-87.
213. Ishiguro, T., et al., *Differential expression of nanog1 and nanogp8 in colon cancer cells*. *Biochem Biophys Res Commun*, 2012. **418**(2): p. 199-204.
214. Amsterdam, A., et al., *Differential localization of LGR5 and Nanog in clusters of colon cancer stem cells*. *Acta Histochem*, 2013. **115**(4): p. 320-9.
215. Ghaleb, A.M., et al., *Altered intestinal epithelial homeostasis in mice with intestine-specific deletion of the Kruppel-like factor 4 gene*. *Dev Biol*, 2011. **349**(2): p. 310-20.
216. Shields, J.M., R.J. Christy, and V.W. Yang, *Identification and characterization of a gene encoding a gut-enriched Kruppel-like factor expressed during growth arrest*. *J Biol Chem*, 1996. **271**(33): p. 20009-17.

217. Chen, X., et al., *Kruppel-like factor 4 (gut-enriched Kruppel-like factor) inhibits cell proliferation by blocking G1/S progression of the cell cycle*. J Biol Chem, 2001. **276**(32): p. 30423-8.
218. Chen, X., et al., *Transcriptional profiling of Kruppel-like factor 4 reveals a function in cell cycle regulation and epithelial differentiation*. J Mol Biol, 2003. **326**(3): p. 665-77.
219. Zhao, W., et al., *Identification of Kruppel-like factor 4 as a potential tumor suppressor gene in colorectal cancer*. Oncogene, 2004. **23**(2): p. 395-402.
220. Patel, N.V., et al., *Expression of the tumor suppressor Kruppel-like factor 4 as a prognostic predictor for colon cancer*. Cancer Epidemiol Biomarkers Prev, 2010. **19**(10): p. 2631-8.
221. Hu, R., et al., *KLF4 Expression Correlates with the Degree of Differentiation in Colorectal Cancer*. Gut Liver, 2011. **5**(2): p. 154-9.
222. Dang, D.T., et al., *Overexpression of Kruppel-like factor 4 in the human colon cancer cell line RKO leads to reduced tumorigenicity*. Oncogene, 2003. **22**(22): p. 3424-30.
223. Rowland, B.D., R. Bernards, and D.S. Peeper, *The KLF4 tumour suppressor is a transcriptional repressor of p53 that acts as a context-dependent oncogene*. Nat Cell Biol, 2005. **7**(11): p. 1074-82.
224. Leng, Z., et al., *Kruppel-like factor 4 acts as an oncogene in colon cancer stem cell-enriched spheroid cells*. PLoS One, 2013. **8**(2): p. e56082.
225. Knutsen, T., et al., *Definitive molecular cytogenetic characterization of 15 colorectal cancer cell lines*. Genes Chromosomes Cancer, 2010. **49**(3): p. 204-23.
226. Wei, D., et al., *KLF4alpha up-regulation promotes cell cycle progression and reduces survival time of patients with pancreatic cancer*. Gastroenterology, 2010. **139**(6): p. 2135-45.
227. Liu, Y., et al., *Demethylation of CpG islands in the 5' upstream regions mediates the expression of the human testis-specific gene MAGEB16 and its mouse homolog Mageb16*. BMB Rep, 2014. **47**(2): p. 86-91.
228. Saitou, M., S.C. Barton, and M.A. Surani, *A molecular programme for the specification of germ cell fate in mice*. Nature, 2002. **418**(6895): p. 293-300.
229. Payer, B., et al., *Generation of stella-GFP transgenic mice: a novel tool to study germ cell development*. Genesis, 2006. **44**(2): p. 75-83.
230. Chen, H.H., et al., *DAZL limits pluripotency, differentiation, and apoptosis in developing primordial germ cells*. Stem Cell Reports, 2014. **3**(5): p. 892-904.
231. Lin, Y. and D.C. Page, *Dazl deficiency leads to embryonic arrest of germ cell development in XY C57BL/6 mice*. Dev Biol, 2005. **288**(2): p. 309-16.
232. Ballow, D.J., et al., *Sohlh2 is a germ cell-specific bHLH transcription factor*. Gene Expr Patterns, 2006. **6**(8): p. 1014-8.
233. Hao, J., et al., *Sohlh2 knockout mice are male-sterile because of degeneration of differentiating type A spermatogonia*. Stem Cells, 2008. **26**(6): p. 1587-97.
234. Campolo, F., et al., *Essential role of Sox2 for the establishment and maintenance of the germ cell line*. Stem Cells, 2013. **31**(7): p. 1408-21.
235. Okamura, D., et al., *Requirement of Oct3/4 function for germ cell specification*. Dev Biol, 2008. **317**(2): p. 576-84.

- 236. Yamaguchi, S., et al., *Nanog expression in mouse germ cell development*. Gene Expr Patterns, 2005. **5**(5): p. 639-46.
- 237. Murakami, K., et al., *NANOG alone induces germ cells in primed epiblast in vitro by activation of enhancers*. Nature, 2016. **529**(7586): p. 403-407.
- 238. Hayashi, K. and M.A. Surani, *Self-renewing epiblast stem cells exhibit continual delineation of germ cells with epigenetic reprogramming in vitro*. Development, 2009. **136**(21): p. 3549-56.
- 239. Chuma, S., et al., *Spermatogenesis from epiblast and primordial germ cells following transplantation into postnatal mouse testis*. Development, 2005. **132**(1): p. 117-22.
- 240. Ikawa, M., et al., *The putative chaperone calmeglin is required for sperm fertility*. Nature, 1997. **387**(6633): p. 607-11.
- 241. Zhang, Z., et al., *MEIG1 is essential for spermiogenesis in mice*. Proc Natl Acad Sci U S A, 2009. **106**(40): p. 17055-60.
- 242. Kin, K., et al., *The Transcriptomic Evolution of Mammalian Pregnancy: Gene Expression Innovations in Endometrial Stromal Fibroblasts*. Genome Biol Evol, 2016. **8**(8): p. 2459-73.

Anexo I: Figuras y tablas

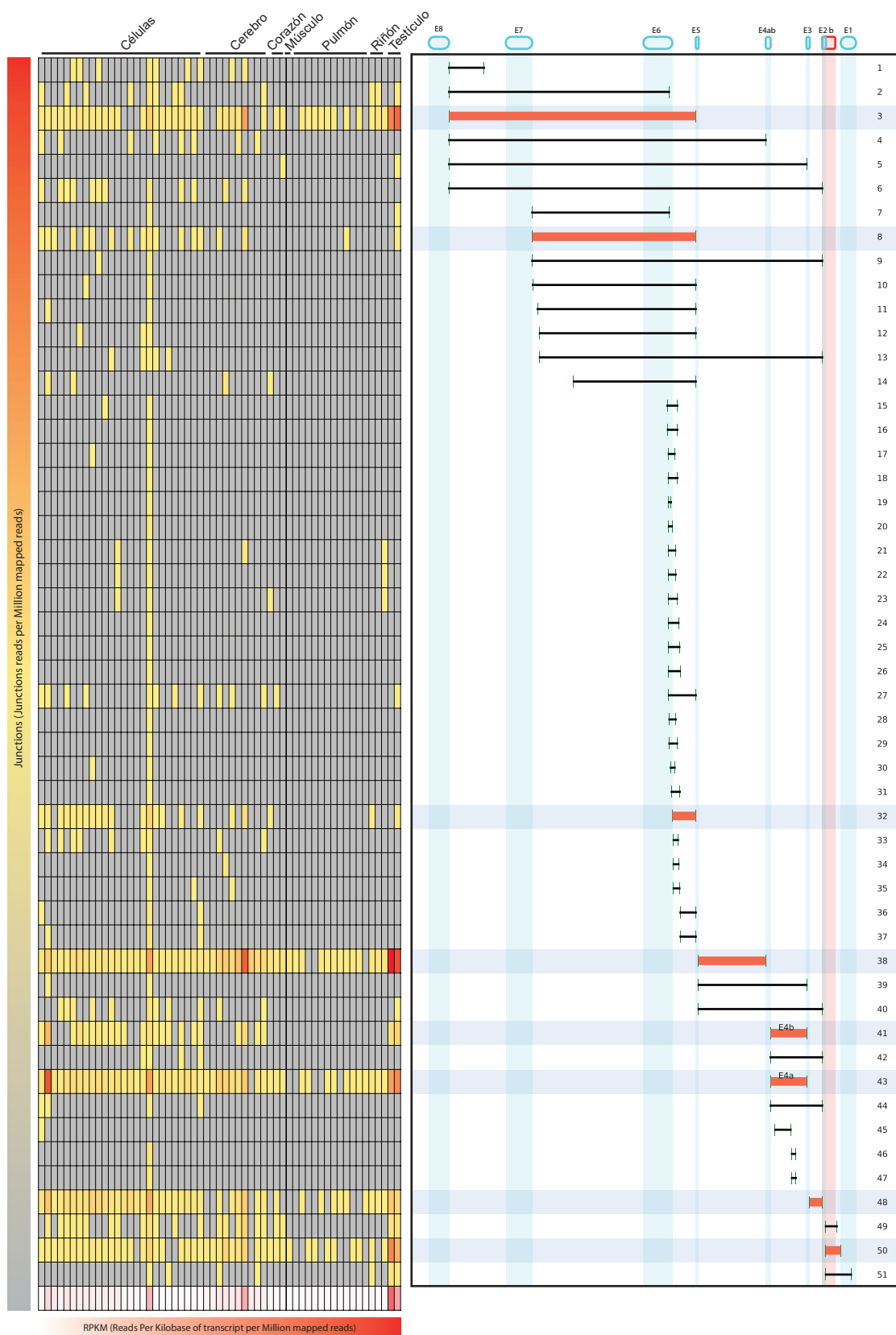


Figura S1. Uniones de exones encontradas en los experimentos de RNA-Seq de humano.



Figura S2. Uniones de exones encontradas en los experimentos de RNA-Seq de ratón.

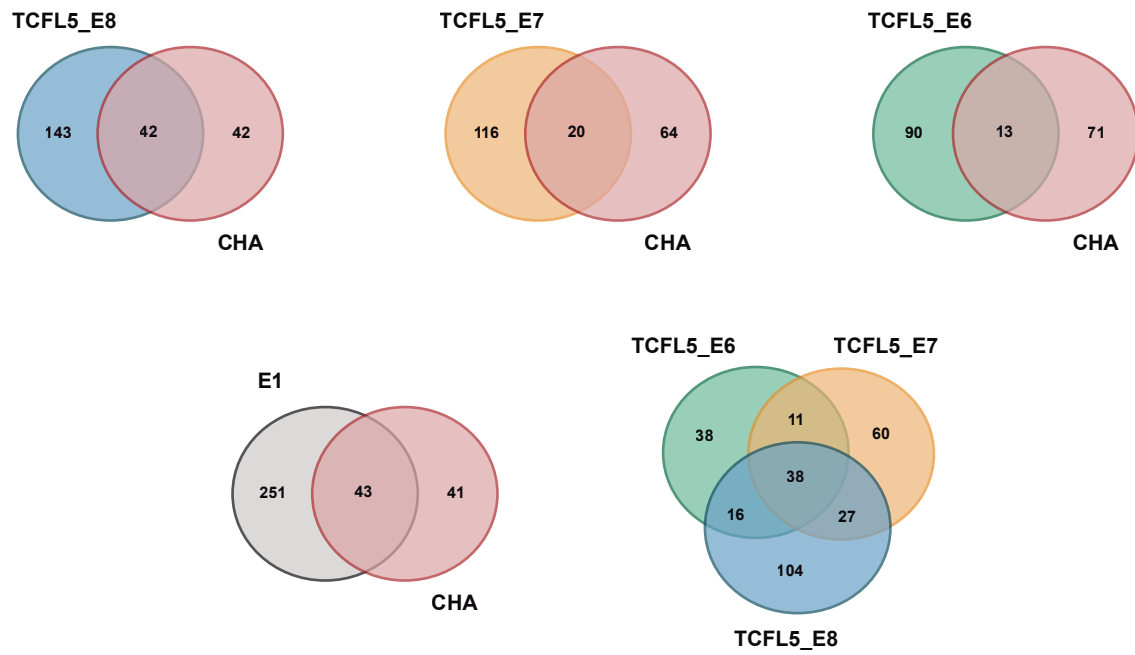


Figura S3. Comparación de las proteínas que interaccionan con cada una de las isoformas. Se muestran distintos diagramas de Venn comparando el número de proteínas de interacción para las isoformas TCFL5_E8, TCFL5_E7, TCFL5_E6 y CHA obtenidas de los resultados del interactoma.

Anexo I: Figuras y tablas

Muestra	Tipo de lecturas	Lecturas totales	Lecturas alineadas	PMID	SRA Id	Descripción
A549	2x75	70.457.234	133.257.238	Encode project	SRR307907 SRR307908	Células epiteliales derivadas de carcinoma de pulmón
K562	1x75	62.732.735	44.264.804	Encode project	SRR521468 SRR534174 SRR534175	Granulocitos indiferenciados derivados de leucemia
Jurkat	1x36	45.756.579	29.041.789	Encode project	SRR492030 SRR492031	Células T derivadas de linfoma
Célula T + PMA	1x50	46.251.043	49.354.842	22454538	SRR408754 SRR408755 SRR408756	Células JSL-1 activados con PMA
Célula T 1	1x50	47.881.096	53.412.582	22454538	SRR408751 SRR408752 SRR408753	Células JSL-1 en reposo
Célula T 2	1x75	30.932.008	24.870.321	21933919	SRR497705	SUP-T1 derivadas de linfoma
Célula T 3	1x75	32.330.868	26.819.184	21933919	SRR497706	SUP-T1 derivadas de linfoma
Célula T 4	1x75	30.121.789	20.078.000	21933919	SRR497707	SUP-T1 derivadas de linfoma
Célula T 5	1x75	32.383.846	30.268.430	21933919	SRR497708	SUP-T1 derivadas de linfoma
Célula T 6	1x75	25.428.142	19.737.092	21933919	SRR497709	SUP-T1 derivadas de linfoma
Célula T+ VIH 1	1x75	31.362.049	20.160.072	21933919	SRR497699	SUP-T1 derivadas de linfoma infectadas con VIH
Célula T+ VIH 2	1x75	35.603.888	23.993.255	21933919	SRR497700	SUP-T1 derivadas de linfoma infectadas con VIH
Célula T+ VIH 3	1x75	29.770.516	12.974.856	21933919	SRR497701	SUP-T1 derivadas de linfoma infectadas con VIH
Célula T+ VIH 4	1x75	23.936.312	13.200.351	21933919	SRR497702	SUP-T1 derivadas de linfoma infectadas con VIH
Célula T+ VIH 5	1x75	24.419.121	12.731.631	21933919	SRR497703	SUP-T1 derivadas de linfoma infectadas con VIH
Célula T+ VIH 6	1x75	23.372.815	13.079.882	21933919	SRR497704	SUP-T1 derivadas de linfoma infectadas con VIH
PBMCs 3	2x101	236.532.985	367.740.205	Encode project	SRR545689 SRR545690	PBMCs obtenidos de donante sano
PBMCs + LPS	2x101	83.210.901	79.723.569	No publicado	SRR629558	PBMCs tratados con LPS
Monocitos 1	2x77	202.993.398	396.413.943	Encode project	SRR545695 SRR545696 SRR545697 SRR545698 SRR545699 SRR545700	Monocitos (CD14+)
Monocitos 2	2x101	36.443.544	18.826.146	23029029	SRR452328	Monocitos M1
Monocitos 3	2x101	30.952.877	27.895.713	23029029	SRR452329	Monocitos M1
Monocitos 4	2x101	22.072.388	13.649.664	23029029	SRR452330	Monocitos M1
Monocitos 5	2x101	51.417.553	51.452.790	23029029	SRR452331	Monocitos M2
Monocitos 6	2x101	15.585.860	12.525.415	23029029	SRR452332	Monocitos M2
Monocitos 7	2x101	49.258.388	38.807.954	23029029	SRR452333	Monocitos M2
Célula B	2x77	169.661.447	322.655.255	Encode project	SRR534319 SRR534320 SRR534321 SRR534322 SRR534323 SRR534324	Células B (CD20+)
Cerebro 1	1x32	17.246.957	17.932.079	18978789	SRR036966	Cerebro
Cerebro 2	1x32	31.940.303	32.349.055	18978789	SRR014262	Córtex Cerebral
Cerebro 3	1x75	24.513.415	18.202.098	23258890	SRR306838	Cerebro
Cerebro 4	1x75	22.576.705	2.336.081	23258890	SRR306840	Cerebro
Cerebro 5	1x75	24.325.223	15.500.956	23258890	SRR306841	Cerebro

Cerebro 6	1x75	18.850.030	10.792.955	23258890	SRR306839	Cerebro
Cerebro 7	1x75	17.422.994	3.519.874	23258890	SRR306842	Cerebro
Cerebro 8	1x75	7.913.181	2.491.510	23258890	SRR306843	Cerebro
Cerebro 9	1x75	32,698,558	22.803.117	23258890	SRR306844	Cerebelo
Cerebro 10	1x75	46.755.221	27.686.265	23258890	SRR306845 SRR306846	Cerebelo
Corazon 1	1x75	24.128.204	17.703.319	23258890	SRR306847	Corazón
Corazon 2	1x75	30.896.351	18.212.717	23258890	SRR306848 SRR306849	Corazón
Corazon 3	1x75	25.197.713	14.621.873	23258890	SRR306850	Corazón
Músculo	1x32	22.640.454	24.302.639	18978789	SRR014266	Músculo esquelético
Pulmón 1	1x32	25.862.057	27.155.165	18978789	SRR014265	Pulmón
Pulmón 2	1x43	13.025.469	14.322.935	22375630	SRR094912	Pulmón fumador
Pulmón 3	1x43	19.179.727	19.062.028	22375630	SRR094913	Pulmón fumador
Pulmón 4	1x43	16.615.218	16.981.614	22375630	SRR094911	Pulmón fumador
Pulmón 5	1x43	14.204.752	13.128.969	22375630	SRR094910	Pulmón fumador
Pulmón 6	1x43	15.201.510	16.238.322	22375630	SRR094909	Pulmón fumador
Pulmón 7	1x43	15.837.368	16.447.042	22375630	SRR094908	Pulmón fumador
Pulmón 8	1x43	17.670.811	19.571.068	22375630	SRR094907	Pulmón no fumador
Pulmón 9	1x43	19.788.339	20.255.816	22375630	SRR094906	Pulmón no fumador
Pulmón 10	1x43	15.886.268	17.166.022	22375630	SRR094864	Pulmón no fumador
Pulmón 11	1x43	17.193.664	18.376.019	22375630	SRR094862	Pulmón no fumador
Pulmón 12	1x43	17.356.854	18.381.134	22375630	SRR094863	Pulmón no fumador
Riñón 1	1x75	22.493.518	16.399.162	23258890	SRR306851	Riñón
Riñón 2	1x75	20.684.752	14.615.495	23258890	SRR306852	Riñón
Riñón 3	1x75	31.386.619	19.274.749	23258890	SRR306853	Riñón
Testículo 1	1x75	9.224.054	4.279.539	23258890	SRR306857	Testículo
Testículo 2	1x75	32.444.809	22.483.468	23258890	SRR306858	Testículo

Tabla S1. Información de los datos de RNA-Seq humanos utilizados para el análisis de la expresión de *TCFL5* y descripción de sus isoformas.

Región	Factores de transcripción									
5' upstream E1	ARID3A	ATF1	BHLHE40	CBX3	CCNT2	CEBPB	CHD2	CREB1	CTCF	E2F6
	EBF1	EGR1	EP300	FOXP2	GABPA	GATA2	GTF2F1	HDAC1	HDAC2	HMGN3
	HNF4A	IRF1	JUND	KDM5B	MAX	MAZ	MXI1	MYBL2	Myc	NR2F2
	NRF1	PAX5	PHF8	POLR2A	RBBP5	RCOR1	REST	SAP30	SIN3AK20	SIRT6
	SMC3	SP1	STAT5A	TAF1	TAL1	TBL1XR1	TBP	TEAD4	TEAD5	TRIM28A
E1 e II	USF1	USF2	YY1							
	SIN3AK20 ZNF263	MXI1 PHF8	MYC RCOR1	CCNT2 MAX	CTCF E2F6	RAD21 GABPA	YY1 FOXP2	E2F1 SMC3	TAF1 POLR2A	NRF1

Tabla S2. Factores de transcripción que se unen al promotor de *TCFL5* analizados por CHIP-Seq en el proyecto ENCODE.

Muestra	Tipo de lecturas	Lecturas totales	Lecturas alineadas	PMID	SRA Id	Descripción
Bazo 1	2x50	114.072.257	255.432.575	23258891	SRR594400	Bazo
Bazo 2	2x80	114.814.142	214.692.757	23258891	SRR594408	Bazo
Bazo 3	2x50	113.321.013	250.239.575	23258891	SRR594417	Bazo
Cerebro 1	1x75	76.170.802	53.064.101	23258890	SRR306758	Cerebro
					SRR306759	
					SRR306760	
					SRR306761	
Cerebro 2	1x75	19.726.026	9.550.600	23258890	SRR306762	Cerebro
Cerebro 3	1x75	41.340.785	24.268.078	23258890	SRR306763	Cerebro
Cerebro 4	2x50	87.264.604	177.016.239	23258891	SRR594393	Cerebro
Cerebro 5	1x75	20.947.739	10.693.287	23258890	SRR306764	Cerebelo
Cerebro 6	1x75	28.710.250	17.900.601	23258890	SRR306765	Cerebelo
Cerebro 7	2x80	118.824.353	200.364.582	23258891	SRR594402	Cerebro
Cerebro 8	2x80	32.511.234	66.218.941	23258891	SRR594410	Cerebro
Cerebro 9	1x75	25.094.445	19.201.472	23258890	SRR306757	Cerebro
Corazón 1	1x75	44.668.984	25.008.946	23258890	SRR306766	Corazón
Corazón 2	1x75	24.493.681	16.469.930	23258890	SRR306767	Corazón
Corazón 3	1x75	25.903.961	18.154.201	23258890	SRR306768	Corazón
Corazón 4	2x36	35.175.982	78.095.213	23258891	SRR594395	Corazón
Corazón 5	2x40	15.968.605	33.571.300	23258891	SRR594412	Corazón
Hígado 1	1x75	48.306.727	22.309.280	23258890	SRR306772	Hígado
Hígado 2	2x50	116.292.478	299.833.542	23258891	SRR594397	Hígado
Hígado 3	1x75	18.444.416	17.069.549	23258890	SRR306773	Hígado
Hígado 4	2x80	134.045.721	208.430.656	23258891	SRR594405	Hígado
Hígado 5	1x75	34.010.208	29.932.099	23258890	SRR306774	Hígado
Hígado 6	2x40	34.824.609	104.527.412	23258891	SRR594414	Hígado
Pulmón 1	2x36	34.050.626	74.914.430	23258891	SRR594398	Pulmón
Pulmón 2	2x80	62.362.901	100.665.712	23258891	SRR594406	Pulmón
Pulmón 3	2x50	111.879.091	232.907.600	23258891	SRR594415	Pulmón
Riñón 1	1x75	45.613.332	15.921.231	23258890	SRR306769	Riñón
Riñón 2	1x75	23.639.764	15.974.541	23258890	SRR306770	Riñón
Riñón 3	1x75	29.158.234	18.200.720	23258890	SRR306771	Riñón
Riñón 4	2x50	119.274.786	249.873.082	23258891	SRR594396	Riñón
Riñón 5	2x75	118.885.190	231.393.521	23258891	SRR594404	Riñón
Riñón 6	2x36	29.821.800	79.881.989	23258891	SRR594413	Riñón
Testículo	1x75	29.762.970	19.245.597	23258890	SRR306776	Testículo

Tabla S3. Información de los datos de RNA-Seq de ratón utilizados para el análisis de la expresión de *Tcf15* y descripción de sus isoformas.